



Yapay Zekâ Sistemleri Testi ve TMMi

#TMMiileBaşarı

EDİTÖR: ERIK VAN VEENENDAAL

www.tmmi.org

TMMi, uygulanan yaşam döngüsünden veya test edilen sistem türünden bağımsız olacak şekilde tasarlanmıştır. Bu doküman, TMMi test iyileştirme modelinin YZ (Yapay Zekâ - YZ) sistemlerini test ederken nasıl kullanılıp faydalı bir şekilde uygulanabileceğini açıklamaktadır. Seviye 2 TMMi süreç alanları ve hedefleri tek tek ele alınmaktadır. Bunların YZ sistemlerini test etmekle nasıl ilişkili olduğu ve uygulamaların genellikle nasıl olacağı belirtilmektedir. YZ sistemlerini test ederken bir uygulamanın katma değeri olmadığı için gerçekleştirilmesi beklenmiyorsa, bu da açıkça belirtilmektedir. Bu dokümanın gelecekteki sürümlerinde, daha yüksek TMMi seviyelerindeki süreç alanları da ele alınacaktır. TMMi seviye 3 ile, esneklik, uyarlanabilirlik, özerklik ve yanlılık gibi YZ tabanlı sistemlere özgü kalite özelliklerinin test edilmesi tartışılacaktır. Bu dokümanın, YZ sistemlerinin test edilmesine ilişkin tam bir müfredat olması amaçlanmadığını, daha çok YZ sistemlerini test ederken TMMi hedeflerinin nasıl yorumlanması gerektiğini “sadece” gösteren ve daha ileri çalışmalar için fikir ve yönlendirmeler sunan bir dokümanın olduğunu unutulmamalıdır. “Yapay Zekâ Sistemleri Testi ve TMMi” dokümanı, tam TMMi modelinin yerini alacak bir açıklama olması da amaçlanmamıştır. YZ sistemlerinin test edilmesi ve TMMi, her zaman TMMi referans modeli dokümanı ile birlikte kullanılmalıdır.

© TMMi Foundation 2011–25

Tüm hakları saklıdır. Bu yayının hiçbir bölümü, ilgili lisans dokümanlarında açıklanan şekilde hariç olmak üzere, TMMi Vakfı'nın önceden izni olmadan herhangi bir şekilde veya herhangi bir yöntemle ödünç verilemez, satılamaz, devredilemez, çoğaltılamaz veya aktarılamaz. İlgili lisans dokümanlarının koşulları uyarınca herhangi bir şekilde kopyalanmasına izin verilen durumlarda, bu bildirim tüm kopyalarda yer alması şartıyla izin verilir.

Ticari marka olduğunu düşündüğümüz kelimeler bu şekilde belirtilmiştir. Ancak, bu tür bir belirtimin varlığı veya yokluğu, herhangi bir ticari markanın yasal statüsünü etkilediği şekilde değerlendirilmemelidir.

TMMi, TMMi Vakfı'nın (İngiltere) tescilli ticari markasıdır.

Katkıda Bulunanlar

Amanda Dean (Australia)

Mattijs Kemmink (The Netherlands)

Vipul Kocher (India)

Poonam Jain (India)

Qin Liu (China)

Srinivas Padmanabhuni (India)

Çiçek Tuna (Germany)

Erik van Veenendaal (The Netherlands)

Kelly Ye (Australia)

Revizyonlar

Bu bölüm, bu dokümanın yayınları arasındaki önemli değişiklikleri özetlemektedir.

Bu bölüm sadece bilgi amaçlıdır.

Sürüm	Düzeltilme Notları
V1.0	"Yapay Zekâ Sistemleri Testi ve TMMi" dokümanının ilk sürümü.

Teşekkür

Yapay Zekâ Sistemleri Testi ve TMMi® dokümanının Türkçeleştirme çalışmasına katkıda bulunan Yazılım Test ve Kalite Derneği çeviri çalışma grubu üyelerine burada tekrar teşekkür etmek isteriz. Çalışma grubu üyeleri (alfabetik sıraya göre):

- Deniz Onat
- Eda Civar
- L. Koray Yitmen
- Meltem Aydaş
- Özlemnur Bayram Dönmez
- Yazılım Test ve Kalite Derneği, Ocak 2026

Yazılım Test ve Kalite Derneği Hakkında

(Turkish Testing Board – www.turkishtestingboard.org) Yazılım Test ve Kalite Derneği, 2006 yılından bu yana Türkiye bilişim sektöründe yazılım testi farkındalığının artması ve gelişmesi için kâr amacı gütmeyen gönüllü bir şekilde aşağıdaki faaliyetleri gerçekleştirmektedir:

Uluslararası Sertifikasyon Sınavları

Dernek uluslararası ISTQB® sertifikasyon sınavlarını gerçekleştirerek sınavlarda başarılı olan katılımcılara uluslararası geçerliliği olan sertifikalar vermektedir. 2006 yılından bu yana 7.000'den fazla test uzmanı adayını derneğe başvurarak sertifikasyon sınavlarına girmiştir. Dernek bünyesinde düzenlenmekte olan sertifikasyon sınavları:

- ISTQB® Uluslararası Sertifikalı Temel Seviye Yazılım Test Uzmanı Sınavı
- ISTQB® Uluslararası Sertifikalı Temel Seviye Çevik Test Uzmanı Sınavı
- ISTQB® Uluslararası Sertifikalı İleri Seviye – Test Analisti Sınavı
- ISTQB® Uluslararası Sertifikalı İleri Seviye – Teknik Test Analisti Sınavı
- ISTQB® Uluslararası Sertifikalı İleri Seviye – Test Yönetimi Sınavı
- ISTQB® Uluslararası Sertifikalı İleri Seviye – Test Otomasyon Mühendisliği Sınavı
- ISTQB® Uluslararası Sertifikalı Performans Testi Sınavı
- ISTQB® Uluslararası Sertifikalı Yapay Zeka Testi Sınavı
- ISTQB® Uluslararası Sertifikalı Mobil Uygulama Testi Sınavı
- ISTQB® Uluslararası Sertifikalı Otomotiv Yazılımı Testi Sınavı
- ISTQB® Uluslararası Sertifikalı Oyun Testi Sınavı
- ISTQB® Uluslararası Sertifikalı Üretken Yapay Zeka ile Test Sınavı

Uluslararası Testİstanbul Konferansları – www.testistanbul.org

Dernek, 2010 yılından bu yana Uluslararası Testİstanbul Konferanslarını düzenlemektedir. Geçtiğimiz 16 konferansta 40'dan fazla ülkeden, 80'den fazla konuşmacı ve 7.400'den fazla katılımcı ağırlanmıştır.

Paneller

Dernek, yazılım test sektörünün gelişimi için sektör veya konu bazlı paneller organize etmektedir. Bu panellere şu ana kadar 1.000'den fazla profesyonel katılım göstermiştir. Şimdiye kadar düzenlenen paneller:

- TestFintech,
- TestDefence,
- TestGames,
- TestInsurance,
- TestAnkara,
- Testİzmir,
- Test Finance

Turkey Software Quality Report (TSQR)

Dernek tarafından 2011 yılından itibaren yüzlerce bilişim profesyoneli ve akademisyeninin katılımıyla her yıl düzenlenen anket sonuçlarının değerlendirilmesiyle hazırlanan, Türkiye bilişim sektörüne yön verir nitelikte çıkarımların olduğu rapordur. İngilizce yayınlanan rapor tüm ISTQB® üye dernekleri aracılığıyla 100'den fazla ülkedeki bilişim profesyoneline ulaşmaktadır.

ISTQB® Worldwide Software Testing Practices Report (WSTPR)

ISTQB® tarafından 100'den fazla ülkeden binlerce bilişim profesyoneli ve akademisyeninin katılımıyla düzenlenen anket sonuçlarının değerlendirilmesiyle hazırlanan, dünya bilişim sektörüne yön verir nitelikte çıkarımların olduğu rapordur.

Türkçeleştirme Çalışmaları

Uluslararası yazılım test terminolojisinin ülkemize kazandırılması için dernek bünyesinde yer alan gönüllü çeviri grubu ISTQB® dokümanlarının çevirisi üzerinde çalışmaktadır. Şu ana kadar çevrilen dokümanlar:

- ISTQB® Uluslararası Sertifikalı Temel Seviye Yazılım Test Uzmanı Ders Programı v4.0
- ISTQB® Uluslararası Sertifikalı Temel Seviye Yazılım Test Uzmanı Örnek Sınav A v4.0
- ISTQB® Uluslararası Sertifikalı Temel Seviye Yazılım Test Uzmanı Örnek Sınav B v4.0
- ISTQB® Uluslararası Sertifikalı Temel Seviye Yazılım Test Uzmanı Örnek Sınav C v4.0
- ISTQB® Uluslararası Sertifikalı Temel Seviye Yazılım Test Uzmanı Ders Programı 2018
- ISTQB® Uluslararası Sertifikalı Temel Seviye Yazılım Test Uzmanı Ders Programı 2011
- ISTQB® Uluslararası Sertifikalı İleri Seviye – Test Analisti Ders Programı 2012
- ISTQB® Uluslararası Sertifikalı Mobil Uygulama Testi Ders Programı 2019
- ISTQB® Uluslararası Sertifikalı Yapay Zekâ Testi Ders Programı 2021
- ISTQB® Uluslararası Sertifikalı Otomotiv Yazılımı Test Uzmanı Ders Programı 2018
- ISTQB® Yazılım Testi Terimler Sözlüğü
- Bir Ejderhadan Yazılım Test Dersi • Çevik Bir Dünyada TMMi®
- TMMi® Genel Broşürü • TMMi® Lightning Scan Tool dokümanı
- TMMi® Kariyer Yolu Broşürü • TMMi® TPI Broşürü
- TMMi® Professional Sertifikasyon Broşürü • DevOps Dünyasında TMMi®

Burslar

Dernek her yıl karından belli bir miktarı T.C. Üniversitelerinin Bilgisayar/Yazılım Mühendisliği ve Bilgisayar Programcılığı, Yönetim Bilişim Sistemleri ve bununla alakalı bölümlerinde okumakta olan başarılı ve ihtiyaç sahibi öğrencilere burs olarak aktarmaktadır. 2025 yılı itibariyle burs sağlanan toplam bursiyer sayısı 100'ü geçmiştir.

İçindekiler

Katkıda Bulunanlar	2
Revizyonlar	3
Teşekkür	4
Yazılım Test ve Kalite Derneği Hakkında	5
İçindekiler	6
1. TMMi Kullanarak Yapay Zekâ Tabanlı Sistemlerin Test Edilmesi	7
1.1 Yapay Zekâ	7
1.2 Test Olgunluk Modeli Entegrasyonu (TMMi)	7
1.3 Yapay Zekâ Tabanlı Sistemlerin Test Edilmesi ve TMMi	8
1.3.1 Yapay Zekâ Tabanlı Sistemlerin Test Edilmesi ile Test Süreçlerinde Yapay Zekâ Kullanımı	8
1.3.2 Yapay Zekâ Sistemlerinin Test Edilmesindeki Zorluklar	9
1.3.3 Amaçlar ve Yaklaşım	10
2. TMMi Seviye 2 - Yönetilen	11
2.1 Süreç Alanı – Test Politikası ve Stratejisi (TPS)	11
2.1.1 SG1 Test Politikasının Oluşturulması	11
2.1.2 SG2 Test Stratejisinin Oluşturulması	12
2.1.3 SG3 Test Performans Göstergelerinin Oluşturulması	13
2.2 Süreç Alanı 2.2 Test Planlama (TPL)	14
2.2.1 SG1 Risk Değerlendirmesinin Yapılması	14
2.2.2 SG2 Test Yaklaşımının Oluşturulması	15
2.2.3 SG3 Test Tahminlerinin Oluşturulması	16
2.2.4 SG4 Test Planının Geliştirilmesi	17
2.2.5 SG5 Test Planına Bağlılık Sağlanması	18
2.3 Süreç Alanı 2.3 Test İzleme ve Kontrol (TMC)	18
2.3.1 SG1 Test İlerlemesinin Plan ile Karşılaştırılarak İzlenmesi	18
2.3.2 SG2 Ürün Kalitesinin Plan ve Beklentilere Göre İzlenmesi	19
2.3.3 SG3 Düzeltici Faaliyetlerin Kapanışa Kadar Yönetilmesi	19
2.4 Süreç Alanı 2.4 Test Tasarımı ve Yürütme (TDE)	19
2.4.1 SG1 Test Tasarım Tekniklerini Kullanarak Test Analizi ve Tasarımı Gerçekleştirme	19
2.4.2 SG2 Test Uygulamasını Gerçekleştirme	20
2.4.3 SG3 Test Yürütme	21
2.4.4 SG4 Test Olaylarını Kapanışa Kadar Yönetme	22
2.5 Süreç Alanı 2.5 Test Ortamı (TEV)	22
2.5.1 SG1 Test Ortamı Gereksinimlerini Geliştirme	22
2.5.2 SG2 Test Ortamı Uygulamasını Gerçekleştirme	22
2.5.3 SG3 Test Ortamlarını Yönetme ve Kontrol Etme	23
Referanslar	24

1 TMMi Kullanarak Yapay Zekâ Tabanlı Sistemlerin Test Edilmesi

1.1 Yapay Zekâ

Yapay zekâ (YZ), en dönüştürücü teknolojik devrimlerden birinin ön saflarında yer almaktadır. Bilgisayar biliminin bir dalı olan yapay zekâ, bir zamanlar sadece insan zihnine özgü olan düşünme, öğrenme ve uyum sağlama yeteneklerini makinelere kazandırmayı amaçlamaktadır. Endüstrileri, toplumu ve günlük hayatımızı yeniden şekillendirme potansiyeli ile yapay zekâ, birçok çığır açan yeniliğin arkasındaki itici güç haline gelmiştir. Temelinde, yapay zekâ, genellikle insan zekâsı gerektiren görevleri yerine getirebilen bilgisayar sistemlerinin geliştirilmesini kapsar. Bu görevler, basit hesaplamalardan karmaşık problem çözüme, örüntü tanıma, dil anlama ve karar vermeye kadar uzanır. Yapay zekâ sistemleri, büyük miktarda veriyi işleyebilir, örüntüleri tanıyabilir ve deneyimlere dayalı olarak performanslarını sürekli olarak iyileştirebilir, bu da onları hızla veriye dayalı içgörülerle yönlendirilen bir dünyada güçlü araçlar haline getirir.

Yapay Zekâ'nın bir alt kümesi olan makine öğrenimi (MÖ), yapay zekâ sistemlerinin önemli bir parçasıdır. MÖ, algoritmaların geniş veri kümeleri üzerinde eğitilmesini içerir ve algoritmaların açık bir programlama olmadan kalıpları tanımasını ve tahminlerde bulunmasını sağlar. Bu yetenek, veriye dayalı karar vermeyi ve tekrarlayan görevlerin otomasyonunu mümkün kılarak sağlık, finans, pazarlama ve üretim gibi çeşitli alanları değiştirmektedir. Ancak, Yapay Zekâ'nın potansiyeli kendine özgü zorlukları da beraberinde getirmektedir. Etik kaygılar, algoritmik yanlılık, veri kalitesi, gizlilik, yorumlanabilirlik ve şeffaflık, YZ sistemlerinde ele alınması gereken önemli konulardır. Yapay Zekâ, meslekler ve toplum üzerindeki etkisiyle ilgili önemli sorular da gündeme getirmektedir.

1.2 Test Olgunluk Modeli Entegrasyonu (TMMi)

TMMi çerçevesi, TMMi Vakfı tarafından test süreçlerinin iyileştirilmesi için bir kılavuz ve referans çerçevesi olarak geliştirilmiştir ve test yöneticileri, test mühendisleri, geliştiriciler ve yazılım kalitesi uzmanları için önemli olan konuları ele almaktadır. TMMi'de tanımlandığı şekliyle test, en geniş anlamıyla tüm yazılım ürünü kalitesi ile ilgili faaliyetleri kapsar.

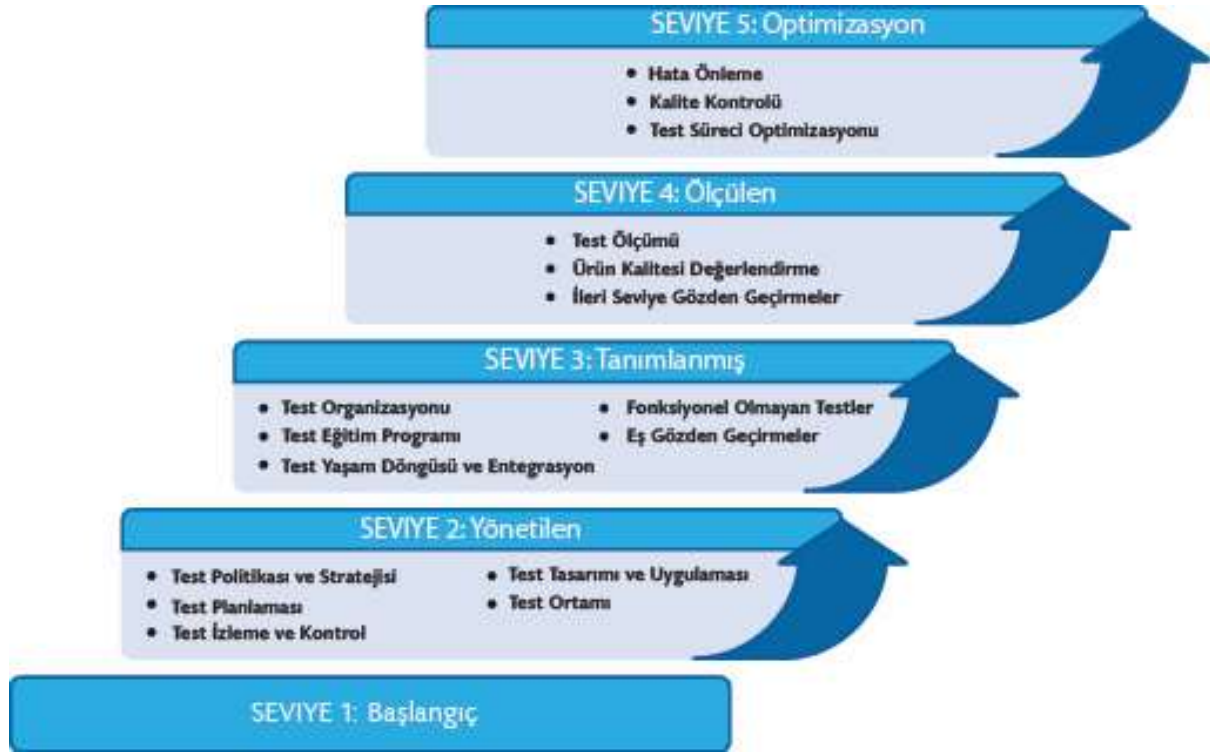
TMMi, süreç değerlendirme ve iyileştirme için olgunluk seviyeleri kavramını kullanır. Ayrıca, süreç alanları, hedefler ve uygulamaları belirler. TMMi olgunluk kriterlerinin uygulanması, test sürecini iyileştirecek ve ürün kalitesi, test mühendisliği verimliliği ve döngü süresi eforu üzerinde olumlu bir etkiye sahip olduğu gösterilmiştir. TMMi, kuruluşların test süreçlerini değerlendirme ve iyileştirme konusunda destek olmak için geliştirilmiştir.

TMMi, süreç iyileştirme için aşamalı bir mimariye sahiptir. Bir kuruluşun test süreci, geçici ve yönetilmeyen bir süreçten yönetilen, tanımlanmış, ölçülen ve optimize edilmiş bir sürece dönüşürken geçirdiği aşamaları veya seviyeleri içerir. Her aşamaya ulaşmak, o aşamanın tüm hedeflerine ulaşıldığını ve iyileştirmelerin bir sonraki aşamanın temelini oluşturduğunu garanti eder.

TMMi'nin iç yapısı, adım adım iyileştirilen kaliteli bir test sürecini desteklemek için sistematik bir şekilde öğrenilebilen ve uygulanabilen zengin test uygulamaları içerir. TMMi'de, olgunluk hiyerarşisini ve test sürecinin iyileştirilmesine yönelik evrimsel yolu belirleyen beş seviye vardır. Her seviye, bir kuruluşun o seviyede olgunluğa ulaşmak için uygulaması gereken bir dizi süreç alanına sahiptir. TMMi'nin her olgunluk seviyesine ait süreç alanları Şekil 1'de gösterilmektedir.

TMMi'nin temel ilkelerinden biri, çeşitli yaşam döngüsü modelleri ve ortamlarına uygulanabilen genel bir model olmasıdır. TMMi tarafından tanımlanan çoğu hedef ve uygulamanın, Çevik (Agile) ve DevOps dahil olmak üzere hem sıralı hem de yinelemeli yaşam döngüsü modellerine uygulanabilir olduğu gösterilmiştir. Ancak, modelin en alt düzeyinde, sağlanan alt uygulamaların ve örneklerin çoğu, uygulanan yaşam döngüsü modeline bağlı olarak (çok) farklıdır. TMMi içinde yalnızca hedeflerin zorunlu olduğunu, uygulamaların zorunlu olmadığı unutulmamalıdır.

TMMi, TMMi Vakfı'nın web sitesinde ücretsiz olarak mevcuttur. Model, İspanyolca, Fransızca ve Çince dahil olmak üzere birçok dile çevrilmiştir. TMMi ayrıca basılı kitap formatında da mevcuttur.



Şekil 1: TMMi olgunluk seviyeleri ve süreç alanları

1.3 Yapay Zekâ Tabanlı Sistemlerin Test Edilmesi ve TMMi

1.3.1 Yapay Zekâ Tabanlı Sistemlerin Test Edilmesi ile Test Süreçlerinde Yapay Zekâ Kullanımı

Yapay zekâ sistemlerinin test edilmesi ile TMMi arasındaki ilişkiyi ve TMMi'nin yapay zekâ sistemlerinin test edilmesi bağlamında nasıl kullanılabileceğini ele almadan önce, "Yapay zekâ sistemlerinin test edilmesi" ile "test süreçlerinde yapay zekâ kullanımı" arasındaki önemli bir farkın vurgulanması gerekmektedir. Bunlar, yazılım test alanında yapay zekânın uygulanmasına ilişkin iki farklı kavramdır.

Yapay zekâ sistemlerinin test edilmesi, yapay zekâ bileşenleri içeren yazılım uygulamalarının veya sistemlerinin performansının, fonksiyonelliğinin ve kalitesinin değerlendirilmesi ve doğrulanması sürecini ifade eder. Temel amaç; bir yazılım sistemi içindeki yapay zekâ algoritmalarının, modellerinin ve işlevlerinin beklendiği şekilde çalışmasını, yeterli sonuçlar üretmesini ve istenen kalite standartlarını karşılamasını sağlamaktır.

Yapay zekâ sistemlerinin test edilmesi; girdi verilerinin test edilmesi, model testleri, entegrasyon testleri, fonksiyonel testler, performans testleri, güvenlik testleri, uyumluluk testleri ile yanlılık (bias) ve tarafsızlık (fairness) gibi etik hususlara yönelik testler dâhil olmak üzere çeşitli test faaliyetlerini kapsamaktadır.

Test süreçlerinde yapay zekâ kullanımı, yazılım test sürecinin kendisini geliştirmek ve optimize etmek amacıyla yapay zekâ teknikleri ve araçlarının uygulanmasını ifade eder. Temel amaç; çeşitli test faaliyetlerini desteklemek ve (kısmen) otomatize etmek, ayrıca veri odaklı kararlar almak için yapay zekâ teknolojilerinden yararlanarak testin verimliliğini, doğruluğunu ve kapsamını artırmaktır. Bununla birlikte yapay zekâ; tahminlerde bulunma, hata tahmini, test analizleri, test senaryosu oluşturma, test verisi oluşturma, testlerin yürütülmesi ve test sonuçlarının analiz edilmesi gibi pek çok test faaliyetinde de kullanılabilir. Ayrıca yapay zekâ, risk ve geçmiş verilere dayalı olarak test senaryolarının önceliklendirilmesine de yardımcı olabilir.

Özetle, “yapay zekâ sistemlerinin test edilmesi”, yapay zekâ ile çalışan yazılımların kalitesinin güvence altına alınmasına odaklanırken; “test süreçlerinde yapay zekâ kullanımı”, test sürecinin kendi verimliliğini ve etkinliğini artırmak amacıyla yapay zekâdan yararlanılmasını ifade eder. Bu doküman kapsamında odak noktası yalnızca yapay zekâ sistemlerinin test edilmesi, bunun TMMi ile ilişkisi ve TMMi’nin bu alanda nasıl katkı sağlayabileceği üzerinedir. Test süreçlerinde yapay zekâ kullanımı ise, TMMi’nin bir sonraki sürümü (TMMi 2.0) için test uygulamaları kapsamında değerlendirilebilecek bir konu olabilir.

1.3.2 Yapay Zekâ Sistemlerinin Test Edilmesindeki Zorluklar

Yapay zekâ sistemlerini nasıl test ederiz? Özellikle en yaygın biçimi olan makine öğrenimi, testlerimize yanıt olarak davranışlarını değiştirebildiği için bu soru daha da karmaşık hâle gelmektedir. Ayrıca testte, bir testin geçtiğini mi yoksa kaldığını mı söyleyebilmek isteriz. Ancak bir yapay zekâ sistemi için “bir testin geçmesi” tam olarak ne anlama gelir? Manuel testlerde, daha az karmaşık muhakeme gerektiren durumlarda bir testin başarılı ya da başarısız olduğunu söylemek mümkün olabilirken, büyük hacimli veriler söz konusu olduğunda bu her zaman mümkün olmayabilir. Net ve eksiksiz bir gereksinim tanımı olmayabilir, doğru davranış zaman içinde değişebilir ya da sistemin bize sunduğu çıktılar dışında, doğru davranışın ne olduğunu dahi bilmiyor olabiliriz.

Yapay zekâ sistemlerinin test edilmesi sürecinde karşılaşılan bazı zorluklar şunlardır:

- Yapay zekâ sistemleri, mevcut yazılımlarla kıyaslandığında dahi olasılık sayısının son derece yüksek olduğu karmaşık problemleri ele almak için kullanılmaktadır.
- Veriler yanlılıklar (bias) içerebilir ve bu yanlılıklar çok kolay bir şekilde yapay zekâ sistemlerine gömülebilir.
- Yapay zekâ tabanlı sistemlerde model kayması (model drift), veri kayması (data drift) ve/veya kavram kayması (concept drift) yaşanabilir.
- Beklenen sonuçları belirleyebilecek test sonucunu bilen (test oracles) yetersizliği veya tamamen eksikliği.
- Yapay zekâ modelleri için kullanılan performans metrikleri, geleneksel sistem performans metriklerinden genellikle oldukça farklıdır ve birçok test uzmanı için yenidir.
- Geleneksel test teknikleri çoğu durumda artık uygulanabilir değildir; bu nedenle test tekniklerini yeniden düşünmemiz gerekmektedir.
- Aynı şekilde test kapsamını (test coverage) nasıl ölçtüğümüzü de yeniden ele almamız gerekmektedir.

Günümüzde yapay zekâ sistemlerinin test edilmesi zorludur ve bu zorlukların tamamı için kolay çözümler bulunmamaktadır. Yapay zekâ tabanlı sistemlerin test edilmesini öğrenmeye başlamak için ilgi çekici bir yaklaşım, AiU United [AiU] ve ISTQB [ISTQB] tarafından geliştirilen sertifikasyon müfredatlarının incelenmesi olabilir. Genel olarak bakıldığında, bir organizasyonun yapay zekâ tabanlı bir sistemi test etmenin zorluklarıyla yüzleşmeye başlamadan önce belirli bir test olgunluk seviyesine sahip olması gerekmektedir.

1.3.3 Amaçlar ve Yaklaşım

Bu doküman, yapay zekâ sistemlerinin test edilmesinin TMMi test süreç iyileştirme modeli ile birleştirilmesinin, daha yüksek bir test olgunluk seviyesine ve nihayetinde iş hedeflerine ulaşmak için olası bir yol olduğunu açıklamaktadır. Doküman, TMMi'nin yapay zekâ sistemlerinin test edilmesi bağlamında nasıl faydalı bir şekilde kullanılabileceğini ve uygulanabileceğini ortaya koymaktadır. Yapay zekâ sistemlerinin test edilmesi sırasında uygulanmak üzere, TMMi uygulamalarını hayata geçiren geleneksel test yaklaşımlarına kanıtlanmış alternatifler sunulmaktadır. Bu doküman, organizasyonun olgunluk yolculuğunda hangi aşamada olduğundan bağımsız olarak, yapay zekâ sistemlerinin test edilmesiyle ilgilenen kuruluşlara TMMi'nin nasıl yardımcı olabileceğini ele almaktadır. Örneğin, organizasyonlar büyüdükçe ve proje baskıları arttıkça sıklıkla gözden kaçan kritik test uygulamalarının hatırlatılmasını sağlamaktadır. Dokümanda, her bir TMMi süreç alanı ve bunlara ait özel hedefler tek tek ele alınmakta; bunların yapay zekâ sistemlerinin test edilmesiyle nasıl ilişkili olduğu ve test uygulamalarının tipik olarak nasıl görüneceği açıklanmaktadır. Eğer bir TMMi uygulamasının, yapay zekâ sistemi test edilirken katma değer sağlamadığı için uygulanması beklenmiyorsa, bu durum da açıkça belirtilmektedir. Yapay zekâ tabanlı sistemlerin test edilmesine yönelik test uygulamaları, tanımlanmış ve TMMi uygulamaları ile hedeflerine eşleştirilmiştir, ancak ayrıntılı olarak açıklanmamıştır. Yapay zekâ sistemlerinin test edilmesine ilişkin test uygulamaları hakkında ek bilgiler sunan, piyasada mevcut iyi kaynaklar bulunmaktadır; örneğin [Smith].

Bir yapay zekâ sisteminin test edilmesi, "sadece" başka bir sistem türünün test edilmesi olmadığından, test uygulamalarındaki değişiklikler ağırlıklı olarak test mühendisliği faaliyetleriyle ilişkili süreç alanlarında (örneğin test planlama, test tasarımı, fonksiyonel olmayan testler vb.) ortaya çıkmaktadır. TMMi modelinde, test mühendisliği süreç alanları özellikle TMMi Seviye 2 (Yönetilen) ve TMMi Seviye 3 (Tanımlı) seviyelerinde yer almaktadır. Bu nedenle bu doküman, yapay zekâ tabanlı sistemlerin test edilmesi bağlamında yalnızca TMMi Seviye 2 (Yönetilen) ve TMMi Seviye 3 (Tanımlanmış) seviyelerini ele almaktadır.

Bu doküman, özgün TMMi modelinden bağımsız olarak kullanılmak üzere tasarlanmamıştır. Yapay zekâ sistemlerinin test edilmesi bağlamında test süreç iyileştirme faaliyetlerini yürüten kişilere, çeşitli TMMi hedef ve uygulamalarının yorumlanması ve anlamlandırılması konusunda rehberlik sağlamayı amaçlamaktadır. Bu doküman, orijinal TMMi modelini tamamlayıcı niteliktedir ve onunla birlikte, destekleyici bir doküman olarak kullanılmalıdır.

2 TMMi Seviye 2 - Yönetilen

2.1 Süreç Alanı – Test Politikası ve Stratejisi (TPS)

Test Politikası ve Stratejisi süreç alanının amacı, test politikasını geliştirmek ve oluşturmak; ayrıca test faaliyetlerinin (örneğin test türleri ve test seviyeleri) açık ve net bir şekilde tanımlandığı, organizasyon genelinde veya program genelinde geçerli bir test stratejisi oluşturmaktır. Test performansını ölçmek, test faaliyetlerinin değerini ortaya koymak ve iyileştirme alanlarını görünür kılmak amacıyla test performans göstergeleri tanımlanır ve kullanıma alınır.

2.1.1 SG1 Test Politikasının Oluşturulması

Test iyileştirme projesine başlayan her kuruluş, öncelikle bir test politikası tanımlamalıdır. Test politikası, kuruluşun test ve test uzmanları ile ilgili genel test hedeflerini, amaçlarını ve stratejik görüşlerini tanımlar. Test politikasının, kuruluşun genel iş (kalite) politikası ile uyumlu olması önemlidir. Test iyileştirmeleri, test (iyileştirme) politikasında dokümente edilmesi gereken net iş hedefleri tarafından yönlendirilmelidir. Bir kuruluş içindeki tüm paydaşlar arasında test ve test hedefleri konusunda ortak bir görüşe ulaşmak için test politikası gereklidir. Bu ortak görüş, kuruluş genelinde test (süreç iyileştirme) faaliyetlerini uyumlu hale getirmek için gereklidir. Test hedeflerinin asla kendi başlarına bir hedef olmaması gerektiğini, bunların çalışan yazılım ve ürün kalitesini sağlamak için daha üst düzey hedeflerden türetildiği unutulmamalıdır.

Yukarıdaki bilgiler, test edilen sistemin türüne bakılmaksızın her organizasyon için geçerlidir. Öncelikle, YZ'yi, getirdiği riskleri ve farklı test yaklaşımlarının gerekli olacağını anlamak önemlidir. Etik, yanlılık, veri ve düzenlemeler gibi bazı konuların tanıtılması veya daha açık bir şekilde ele alınması gerekebilir. Ayrıca, şeffaflık, yorumlanabilirlik ve açıklanabilirlik gibi YZ sistemlerine özgü fonksiyonel olmayan özelliklerin tanıtılması gerekir. Sonuç olarak, test hedefleri ve amaçları ile test ve test süreci tanımına ilişkin temel görüşlerin yeniden gözden geçirilmesi gerekebilir. Bunların çoğu YZ sistemleri için olanlarla büyük ölçüde aynı olsa da bazı farklılıklar olacaktır. Örneğin, hedefler arasında YZ sistemlerinin getirdiği risklerin azaltılmasına ilişkin bir ifade yer alabilir, süreç tanımında YZ sistemlerinin aynı girdiler için farklı sonuçlar verebileceği gerçeğine atıfta bulunulması gerekebilir, entegrasyondan önce bir model test aşaması olabilir. Geleneksel sistemlerde genellikle böyle bir aşama yoktur.

Etik, test uzmanlarının YZ tabanlı sistemlerde verileri nasıl kullandığını dikkate alması gerekebilir, ancak aynı zamanda sistemin kendisinin etik davranışıyla da ilgilidir. Bir YZ sistemi, gerçek dünyada yaşam ve özgürlük üzerinde ciddi karar sonuçları doğurabileceğinden, etik kararlar risk altındadır ve önemlidir. YZ'deki etik kaygılar, sistemin sonuçlarından mı yoksa yalnızca belirli eğitim veri kümelerinden mi öğrendiğine bağlı olarak farklılık gösterecektir.

Veriler, test politikasında dikkate alınması gereken önemli konular olarak yanlılık, gizlilik ve güvenlik gibi sorunları gündeme getirecektir. İyi veri gereksinimleri ve popülasyonu anlamak son derece önemli hale gelmektedir. Örneğin, yüz tanıma sistemleri, sadece uygun bir veri seti ile değil, etkilenen popülasyonu temsil eden verilerle eğitilmeli ve test edilmelidir. Çok uluslu bir ürün, yerel popülasyondan elde edilebilecek verilerden çok daha geniş bir veri yelpazesine test edilmesi anlamına gelir.

Kalite, güvenlik açısından kritik sistemler gibi alanlar dışında, normalde BT sistemlerinin düzenlemeye tabi bir yönü değildir. Ancak, insanlar hakkında kararlar alan yapay zekâ tabanlı sistemlerin kullanımının artması ve insanların bu kararları tam olarak anlayamaması nedeniyle, kuruluşları kalite konusunda yasal olarak kısıtlayan çeşitli düzenlemeler ilerlemekte veya yürürlüğe girmektedir.

2.1.2 SG2 Test Stratejisinin Oluşturulması

Test stratejisi, test politikasını takip eder ve projelerdeki test faaliyetleri için bir başlangıç noktası görevi görür. Test stratejisi genellikle organizasyon genelinde veya program genelinde tanımlanır. Tipik bir test stratejisi, üst düzey bir ürün risk değerlendirmesine dayanır ve gerçekleştirilecek test türleri ve test seviyelerinin bir açıklamasını içerir. Örneğin, her proje içinde entegrasyon testi ve kabul testi yapılacağını belirtmek yeterli değildir. Entegrasyon ve kabul testlerinin ne anlama geldiğini, bunların genellikle nasıl gerçekleştirildiğini ve ana hedeflerinin neler olduğunu tanımlamamız gerekir. Deneyimler, bir test stratejisi tanımlandığında ve uygulandığında, çeşitli test faaliyetleri arasında daha az çakışma olasılığının olduğunu ve bunun da daha verimli bir test sürecine yol açtığını göstermektedir. Ayrıca, çeşitli test türleri ve seviyelerinin test hedefleri ve yaklaşımları uyumlu olduğundan, daha az açık kalması muhtemeldir, bu da daha etkili bir test sürecine ve dolayısıyla daha yüksek ürün kalitesine yol açar.

Test stratejisi, Yapay Zekâ sistemlerinin test edilmesinde de hayati öneme sahip bir dokümandır. YZ sistemlerinin test edilmesi, birçok organizasyon için görece yeni bir alandır; bu nedenle organizasyon genelinde ortak bir test stratejisi, projelere YZ testlerini nasıl yürütecekleri konusunda rehberlik eder ve projelerin “tekerleği yeniden icat etmesini” önler. Bir test stratejisi ayrıca, test faaliyetlerine dahil olan herkesin büyük resmi anlamasını sağlar. Geleneksel sistemlerde test stratejisi; uygulanacak test seviyeleri ve test türleri etrafında yapılandırılır. Bu seviyeler ve türler, birbirleriyle olan ilişkileri dikkate alınarak tanımlanır ve konumlandırılır.

Ancak yapay zekâ sistemlerinin test edilmesinde bu yaklaşımın büyük bir kısmı farklıdır.

YZ test stratejisi, her şeyden önce iki ana fazı birbirinden ayırır: Çevrimdışı test ve Çevrimiçi test. **Çevrimdışı aşamada**, YZ modeli, hedeflerinin başarısını değerlendirmek için doğruluk veya hassasiyet gibi metrikler kullanılarak ayarlanır, doğrulanır ve test edilir. Tipik olarak, makine öğrenimi modeli fonksiyonel davranış açısından test edilir, ancak model için uygun olan bazı fonksiyonel olmayan testler de yapılır. Bu testler arasında eğitim hızı, tahmin hızı, kullanılan bilgi işlem kaynakları, uyarlanabilirlik ve şeffaflık sayılabilir. Model testine ek olarak, çevrimdışı aşama başka bir yeni test düzeyini de içerir: girdi verisi testi. Girdi verisi testi, sistemin eğitim ve tahmin için kullandığı verilerin en yüksek kalitede olmasını sağlamak içindir.

Çevrimiçi (on-line) faz sırasında, eğitilmiş modelin; diğer yapay zekâ ve yapay zekâ olmayan bileşenler ile operasyonel ortamı da kapsayacak şekilde sistemin geri kalanıyla entegrasyonu test edilir. Bu fazda hem fonksiyonel hem de fonksiyonel olmayan testler (örneğin performans testleri) gerçekleştirilir. Çevrimiçi fazda, geleneksel test yaklaşımlarında kullanılan test seviyeleri ve test türleri, test stratejisini yapılandırmak amacıyla kullanılabilir.

- Bileşen testi, kullanıcı arayüzleri ve iletişim bileşenleri gibi YZ olmayan model bileşenlerine uygulanır.
- Bileşen entegrasyon testi, sistem bileşenlerinin birbiriyle doğru şekilde etkileşimde bulunduğunu doğrulamak amacıyla gerçekleştirilir.
- Sistem testi, entegre bileşenlerden oluşan tam sistemin, operasyonel ortama yakın şekilde temsil edilen bir test ortamında hem fonksiyonel hem de fonksiyonel olmayan bakış açılarından beklendiği

gibi çalıştığını doğrulamak için yapılır. Sisteme bağlı olarak bu testler, beklenen operasyonel ortamda gerçekleştirilen saha denemeleri şeklinde de olabilir.

- Kabul testi, tam sistemin müşteri tarafından kabul edilebilir olup olmadığını belirlemek amacıyla

		Tahmin Edilen	
		Negatif (N) -	Pozitif (P) +
Gerçek	Negatif (N) -	Gerçek Negatif (GN)	Yanlış Pozitif (YP) Tip I Hatası
	Pozitif (P) +	Yanlış Negatif (YN) Tip II Hatası	Gerçek Pozitif (DP)

Şekil 1: Karşıklık Matrisi

kullanılır. Ancak bu aşamada ek odak gerektiren bazı hususlar olabilir. Örneğin sistemin insan faaliyetlerini ikame etmek üzere tasarlandığı durumlarda, sistem davranışlarının insan davranışlarıyla karşılaştırılması ve tanımlanmış performans kriterlerine göre değerlendirilmesi gerekebilir.

- Yapay Zekâ sistemlerinin test edilmesinde de geçerli olan bir test seviyesi, sistemin performansını sürekli olarak izlemektir. Yapay zekâ sistemlerinde, sistemin canlı kullanımdaki kalitesini sürekli olarak gözden geçirme ve yeniden değerlendirme ihtiyacı sıklıkla ortaya çıkar. Üretimde kaliteyi izlemenin önemli bir nedeni, yanlışlık tespitidir. Eğitim verileri ile modelin canlı ortamda maruz kaldığı veriler arasında potansiyel bir fark olması nedeniyle, test ortamında yanlışlığı belirlemek bazen zor olabilir ve sürekli izleme bu konuda yardımcı olabilir. Sürekli izlemenin bir diğer önemli nedeni de sapmadır (drift). Yapay zekâ tabanlı sistemlerde model sapması, veri sapması ve kavram sapması, zaman içinde dikkate alınması ve izlenmesi gereken faktörlerdir. Sapma belirtilerini erken tespit etmek için model performans ölçütlerinin sürekli izlenmesi gerekir. Sapma, girdi özellikleri ve çıktı arasındaki ilişkinin değişmesi nedeniyle çalışan modellerin zaman içinde bozulabileceği gerçeğini ifade eder. Model kullanılırken, durumu değişebilir ve model amaçlanan performansından sapabilir.

Yapay zekâ sistemlerinin kuruluşun kendi bünyesinde geliştirilmediği, ancak Microsoft, Apple veya Google gibi dış kaynaklar tarafından geliştirilen Yapay zekâ sistemlerinin (daha büyük bir yapılandırmanın parçası olarak) uygulandığı durumlar da olabilir. Bu durum elbette yapay zekâ test stratejisini de etkileyecektir. Bu strateji, diğer sistemlerle entegrasyona, sistemlerin kuruluşun ihtiyaçlarına göre uyarlanmasına ve yapılandırılmasına odaklanan COTS test stratejisine benzeyecektir.

2.1.3 SG3 Test Performans Göstergelerinin Oluşturulması

Test politikasında tanımlanan test iyileştirme iş hedefleri, bir dizi temel test performans göstergesine dönüştürülmelidir. Test politikası ve beraberindeki performans göstergeleri, beklenen ve elde edilen test performans düzeylerini iletme için net bir yön ve araç sağlar. Performans göstergeleri, testlerin ve test süreci iyileştirmesinin paydaşlar için değerini göstermelidir. Süreç iyileştirmeye yapılan yatırımlar uzun vadeli yönetim desteği gerektirdiğinden, motivasyonlarını korumak için iyileştirme programının faydalarını nicel olarak ölçmek çok önemlidir. Bu TMMi özel hedefinin, sınırlı sayıda (örneğin 2 veya 3) test performans göstergesini tanımlamakla ilgili olduğuna dikkat edilmelidir. Bu, tam bir ölçüm programı oluşturmak ve uygulamakla ilgili değil, testin değerinin zaman içinde ve farklı teslimat ortamlarında nasıl değiştiğini gösteren bir dizi temel gösterge tanımlamakla ilgilidir.

Yapay zekâ sistemlerinin test edilmesi, test süreci iyileştirme programının bir parçasıysa ve düzenli olarak gerçekleştiriliyorsa, test iyileştirme programının sonuçlarını göstermek için yapay zekâ sistemlerinin test edilmesi için özel test performans göstergeleri gerekecektir. Yapay zekâ sistemlerinin

doğası ve buna eşlik eden test hedeflerinin farklılıkları nedeniyle, geleneksel performans göstergeleri (ör. Hata Tespit Yüzdesi) genellikle geçerli değildir. Yapay zekâ sistemleri için test performans göstergeleri, eğitilmiş veya test edilmiş modelin sonuçlarını ne kadar doğru ve güvenilir bir şekilde sunduğunu ölçecektir. Yapay zekâ sistemlerinde kullanılacak performans göstergelerinin, öğrenme türüne (denetimsiz veya denetimli) ve model türüne (kümeleme, ilişkilendirme, sınıflandırma veya regresyon) bağlı olacağına dikkat edilmelidir. Örneğin, denetimli sınıflandırma yapay zekâ sistemleri için bir karışıklık matrisi (bkz. şekil 1) kullanılabilir ve bu doğruluk ve/veya hassasiyet temel alınarak bir performans göstergesi olarak hesaplanabilir ve kullanılabilir.

Denetimli regresyon yapay zekâ sistemlerinde, Ortalama Kare Hatası (MSE) performans göstergesi olarak kullanılabilir [ISTQB]. MSE, gerçek değer ile tahmin edilen değer arasındaki farkların karelerinin ortalamasıdır. MSE değeri her zaman pozitifdir ve sifıra yakın bir değer, daha iyi bir regresyon modelini gösterir. (Test) performans göstergesi olarak kullanılacak diğer metrik örnekleri, AiU Sertifikalı Yapay Zekâ Test Uzmanı müfredatında [AiU] ve ISTQB Sertifikalı Yapay Zekâ Test Uzmanı müfredatında [ISTQB] bulunabilir.

2.2 Süreç Alanı 2.2 Test Planlama (TPL)

Test Planlamasının amacı, belirlenen risklere ve tanımlanan test stratejisine dayalı bir test yaklaşımı tanımlamak ve test faaliyetlerini gerçekleştirmek ve yönetmek için sağlam temellere dayanan planlar oluşturmak ve sürdürmektir.

Başarılı test planlamasının anahtarının, ilgili test planını (“doküman”) tanımlamakta değil, önceden düşünmekte (“faaliyet”) olduğu unutulmamalıdır.

2.2.1 SG1 Risk Değerlendirmesinin Yapılması

Kapsamlı testler yapmak imkansızdır. Bu durum, yapay zekâ sistemlerin test edilmesi için de geçerlidir ve her zaman seçimler yapılmalı ve öncelikler belirlenmelidir. Bu nedenle, TMMi hedefi yapay zekâ tabanlı sistemlerin test edilmesi için de geçerlidir. Test yaklaşımı, yapay zekâ tabanlı sistem için bir risk analizine dayalı olacak ve hem geleneksel testleri hem de yapay zekâ bileşenlerine ve yapay zekâ tabanlı sistemlere özgü özellikleri ele almak için daha özel testleri içerecektir.

Yapay zekâ tabanlı sistemlerde ürün risklerini değerlendirmek için kullanılan tipik teknikler arasında beyin fırtınası, uzman görüşmeleri ve kontrol listeleri yer alır.

- Beyin fırtınası, iletişimin anahtar rol oynadığı bir veya daha fazla toplantı aracılığıyla gerçekleştirilir. Genellikle, bir projede yapay zekâ tabanlı sistemin nasıl çalıştığına dair tam bir resme sahip olan kimse yoktur. Özellikle yapay zekâ ile ilgili geniş bir yelpazede uzmanlık becerilerine sahip çok disiplinli ekipler, projenin hiçbir yönünün çoğunlukta olmadığı bu sürecin kalitesi için çok önemlidir. Ekibin amacı, yapay zekâ tabanlı sistemin ve işleyişinin tüm yönleri hakkında geniş bir yelpazede bakış açıları ve algılar sunmaktır.
- Uzmanlarla yapılan görüşmeler, risk değerlendirme ve belirleme sürecini geliştiren değerli bilgiler ve uzmanlık sağlar. Görüşme tekniklerini kullanarak, kuruluşlar yapay zekâ uzmanlarından doğrudan bilgi toplayabilir ve potansiyel riskleri ve bunların etkilerini daha kapsamlı bir şekilde anlayabilir.

- Risk tanımlama kontrol listeleri, önceki yapay zekâ projelerinin deneyimlerini bir araya getirir. İdeal olarak, liste yapay zekâ tabanlı sistemlerin genel risk alanlarına odaklanmalıdır. Amaç, bazen unutulmuş, sıkça karşılaşılan ve dikkate alınabilecek ürün riskleri ile yapay zekâ tabanlı sistemlerin maruz kalabileceği sorunların türlerine ilişkin ipuçları konusunda hızlı düşünmeyi sağlamaktır.

Yapay Zekâ sistemleri için ürün risklerine örnekler:

- Giriş verilerinin kalitesi çok düşük.
- Eğitim için yetersiz veri hacmi bulunuyor
- Operasyonel veri hattıyla ilgili sorunlar var.
- Modeli geliştirmek için kullanılan MÖ iş akışı tamamen doğru olmayabilir.
- MÖ çerçevesi, algoritma, model, model ayarları ve/veya hiper parametrelerin yanlış seçilmiş.
- Üretim ortamında istenen MÖ fonksiyonel performansı elde edilemiyor
- Kullanıcılar, sunulan MÖ fonksiyonel performansından memnun değil.
- Kendi kendine öğrenme sistemi, kullanıcıların beklediği hizmeti sunmuyor.
- Kullanıcılar, sistemin kararlarını nasıl belirlediğini anlamadıkları için hayal kırıklığına uğruyorlar.
- Çalışma modelleri, girdi özellikleri ile çıktı arasındaki ilişkinin değişmesi nedeniyle zamanla bozulur (genellikle kavram kayması olarak adlandırılır).
- Önceden eğitilmiş modellerin kendi bilinmeyen ve/veya dokümente edilmemiş yanlılıkları ve eksiklikleri vardır.
- Halüsinasyonlar, yetersiz eğitim verilerinden, modelin açıkları makul görünen metinlerle doldurma eğiliminden veya komutları yanlış yorumlamasından kaynaklanabilir ve yapay zekâ genellikle uydurma çıktılarını yüksek bir güvenle sunduğu için önemli bir sorun oluşturabilir.

2.2.2 SG2 Test Yaklaşımının Oluşturulması

Tanımlanan ve önceliklendirilen ürün risklerini azaltmak için bir test yaklaşımı tanımlanır. Test yaklaşımı, tanımlanan kurumsal test stratejisi tarafından da yönlendirilir (bkz. paragraf 2.1.2). İlgili kişilere ne yapmaları gerektiği ve testlerin nasıl gerçekleştirileceği açıkça belirtilmelidir. Test stratejisinde olduğu gibi, ayrıntılı test yaklaşımı da iki ana aşamayı ayırır: uygulamanın çevrimdışı ve çevrimiçi testleri. Gerçekleştirilecek her test seviyesi ve test türü için bir test yaklaşımı tanımlanacaktır.

Çevrimdışı aşama, veri hazırlama, ön işleme ve veri girişi testi (başlangıç test seviyesi) ile başlar. Gerçekleştirilecek faaliyetler arasında veri işleme, veri filtreleme ve veri ön işleme yer alır. Veri ön işleme, veri tamamlama (imputation), yani eksik değerlerin ele alınması, genel bir bakış elde etmek için veri görselleştirme ve anomaliler ile uç değerlerin ele alınması, korelasyon analizi ve boyutların azaltılmasını içerir. Ardından YZ modeli değerlendirilir ve analiz edilir. Bu, mevcut girdi veri setinin eğitim, doğrulama ve test veri setlerine bölünmesini de içerir. Eğitim seti, modelin eğitildiği verileri içerir. Doğrulama veri seti, ML algoritması tarafından eğitimin etkili olup olmadığını değerlendirmek için kullanılır. Eğitim sonrası aşamada, ML modeli test veri seti ile değerlendirilir.

Çevrimiçi aşamada, YZ olmayan bileşenler test edilir ve YZ ile YZ olmayan parçalar arasındaki etkileşim test edilir. Bu, genellikle YZ uygulamasının mimarisinin ayrıntılı bir şekilde incelenmesini ve anlaşılmasını gerektirir. Son olarak, tamamen entegre YZ sistemi test edilir. Keşifsel testlere ek olarak, YZ tabanlı sistemler için test senaryolarını belirlemek ve tanımlamak için popüler bir teknik de metamorfik testtir.

Giriş kriterleri, hazırlık tanımı (özel uygulama 2.3) olarak da adlandırılır ve çıkış kriterleri, tamamlanma tanımı (özel uygulama 2.4) olarak da adlandırılır ve tanımlanmış test yaklaşımının bir parçasıdır. Bir test aşamasının giriş kriterleri genellikle önceki test aşamasının çıkış kriterlerinin yerine getirilmesidir, ancak test hazırlığının tamamlanması ve test ortamının kullanılabilirliği ile de ilgilidir. Test aşamasına bağlı olarak, bazı çıkış kriterleri geleneksel sistemlerdeki benzeşebilir, örneğin test kapsamı ve çözülmemiş hatalar. Geleneksel sistemlerde kullanılan çıkış kriterlerinden çok farklı olan çıkış kriterlerini tanımlamak için temel olarak kullanılacak ölçütler de mevcuttur.

Aşağıdaki tablo, YZ modelinin türüne göre kullanılabilir çıkış kriterleri (metrikler) örneklerini göstermektedir [AiU]:

Öğrenme Tipi	Model Tipi	Kullanılan Çıkış Kriterleri
Denetimsiz	Kümeleme	Eylemsizlik Ayarlanmış Rand Puanı
	Birleşme	Destek Güven Yükseltme
Denetimli	Sınıflandırma	Kesinlik Hassasiyet Hatırlama/Duyarlılık Özgüllük F1 skoru
	Regresyon	Kök ortalama kare hatası R-kare hatası

Tablo 1: Yapay zeka model türüne göre çıkış kriterleri ve ölçütler

2.5 numaralı özel uygulama olan “askıya alma ve yeniden başlatma kriterlerini tanımlama” uygulamasının yapay zekâ tabanlı sistemlerin test edilmesiyle ilgili olup olmadığı, uygulanan yaşam döngüsüne ve benimsenen yaklaşıma bağlıdır. Daha geleneksel bir yaklaşım benimsendiğinde bu uygulama geçerli olacaktır. Çevik ile ise daha az ilgili olacaktır. Çevik ile bu uygulama genellikle durum ve ilerleme toplantılarında dolaylı olarak kullanılır, ancak bunları test aşamaları için önceden açıkça tanımlamak abartılıdır. İlerlemeyi potansiyel veya gerçek bir tehdit olarak değerlendirebilecek engelleyici sorunlar olduğunda, bunlar toplantılarda tartışılır. Toplantıda ekip, sorunları çözmek için alınması gereken önlemleri (varsa) belirler. Çevik proje yönetimi rutini, bu özel uygulama için alternatif bir uygulama görevi görür.

2.2.3 SG3 Test Tahminlerinin Oluşturulması

Test eforu tahmini, bir projenin veya ekibin test hedeflerini karşılamak için gereken testle ilgili iş miktarını tahmin etmeyi içerir. Tahminin bir dizi varsayıma dayandığını ve her zaman tahmin hatasına maruz kaldığını paydaşlara açıkça belirtmek önemlidir. Bu, özellikle bir kuruluşun belirli bir sistem türünü (örneğin, yapay zekâ tabanlı bir sistemi) test etme konusunda daha az deneyime sahip olduğu durumlarda geçerlidir.

Genel olarak, iki tahmin yaklaşımı mevcuttur:

Oranlara dayalı tahmin. Bu metrik tabanlı teknikte, rakamlar kuruluş içindeki önceki projelerden toplanır ve bu da benzer projeler için “standart” oranlar elde edilmesini mümkün kılar. Kuruluşun kendi projelerinin oranları (örneğin, geçmiş verilerden alınan) genellikle tahmin sürecinde kullanılacak en iyi kaynaktır. Bu standart oranlar daha sonra yeni projenin test eforunu tahmin etmek için kullanılabilir. Yapay zekâ test sistemlerinin test edilmesi sektör ve çoğu kuruluş için nispeten yeni olduğundan, bu tahmin yaklaşımını yapay zekâ tabanlı sistemlerin test edilmesinde tercih edilen bir yaklaşım haline getiren çok az (varsa) metrik mevcuttur.

Uzman temelli tahminleme, örneğin Wideband Delphi tekniği. Bu teknikte, “uzmanlar” deneyimlerine dayalı tahminlerde bulunurlar. İşin parçalara ayrılmasına dayalı olarak, her uzman ayrı ayrı bireysel görevler için gereken eforu tahmin eder. Sonuçlar toplanır ve üzerinde anlaşmaya varılan sınırların dışında sapmalar varsa, uzmanlar mevcut tahminlerini tartışır. Planlama Pokeri, Wideband Delphi'nin bir çeşididir ve genellikle Çevik yazılım geliştirmede kullanılır. Metrikler mevcut olmadığından, bu teknik yapay zekâ tabanlı sistemlerin test edilmesiyle ilgili efor ve maliyeti tahmin etmek için tercih edilen teknik gibi görünmektedir. Projeye veya ekibe dahil olanlar ve belki de kuruluş içindeki diğer yapay zekâ projelerine dahil olmuş olanlar tahmin ekibinin bir parçası olabilirler.

Süreci takip ederek, test eforu ve maliyetini tahmin etmek için en iyi tahminlerini yaparlar ve yapılan varsayımları da doküman ederler. Wideband Delphi tahminleme oturumları genellikle, tanımlanmış test yaklaşımına göre geliştirilecek test çalışma ürünlerinin listesi ve test çalışma ürünleriyle ilgili olarak gerçekleştirilecek test görevlerinin listesi gibi girdilerle sağlanır.

Yapay zekâ sistemlerini test etmek için gereken eforu tahmin etmek, genellikle geleneksel yazılım testlerinden daha karmaşıktır. Aşağıda, dikkate alınabilecek yapay zekâya özgü faktörlerin bazı örnekleri verilmiştir:

- Veri karmaşıklığı, Yapay Zekâ sistemleri/uygulamaları için testler, doğru ve tarafsız tahminler sağlamak için veri doğrulama, yanlılık tespiti ve sapma izleme için ekstra efor gerektirir.
- Yapay Zekâ Model Davranışı, geleneksel yazılımlardan farklı olarak, yapay zekâ öğrenir ve gelişir, bu nedenle sürekli test, açıklanabilirlik kontrolleri ve ayarlamaya ihtiyaç duyar.
- Sürekli izleme, yapay zekânın performansını korumak için gerçek zamanlı izleme, performans kontrolleri ve model yeniden eğitimi gerekir.
- Yasalara uygunluk, yapay zekâ testleri etik ve yasal kullanımı sağlamak için GDPR, Yapay Zekâ Yasası ve sektöre özgü kurallara uymalıdır.

2.2.4 SG4 Test Planının Geliştirilmesi

Bir test planı her zaman iki ana taraf göz önünde bulundurularak geliştirilmelidir: (1) test uzmanları (ne yapacaklarını bilmeleri gerekir) ve (2) paydaşlar (gerçekleştirilecek testleri, katılımlarını anlamaları, kaynakları, ilgili riskleri vb. kabul etmeleri gerekir). Yapay zekâ tabanlı bir sistemi test eden test planının ayrıntı düzeyi, uygulanan yaşam döngüsüne de bağlıdır; örneğin, daha çevik bir çalışma şekli, daha hafif bir test planı anlamına gelir.

Yapay zekâ sistemleri için test planı, yapay zekânın veri odaklı, uyarlanabilir ve olasılıksal yapısı nedeniyle geleneksel yazılım testlerinin ötesine geçmelidir. Geleneksel uygulamalardan farklı olarak, yapay zekâ modelleri öğrenir, gelişir ve olasılıksal kararlar alır, bu da özel test stratejileri ve odaklanmış test planları gerektirir. Yapay zekânın karmaşıklığı ve gelişen yapısı göz önüne alındığında, testler sürekli ve otomatikleştirilmiş olmalı ve sürekli doğrulama sağlayan süreçler olmalıdır.

Yapay zekâ tabanlı bir sistemi test etmek birçok kuruluş için nispeten yeni bir konu olduğundan, test uzmanlarının ne yapmaları gerektiğini anlamalarını ve paydaşların iyi bilgilendirilmesini sağlamak için doküman edilmiş bir test planı gerekecektir. Test planı, amacına hizmet ettiği sürece bir doküman, sunum veya zihin haritası formatında olabilir. Test planı, önceki test planlama faaliyetlerinin sonuçlarını, ürün ve proje risklerinin öncelikli bir listesini, hem çevrimdışı aşama (veri girişi testi, model testi) hem de çevrimiçi aşamanın test seviyeleri ve test türleri için yaklaşımı ve faaliyetleri ayrıntılı olarak doküman edecektir.

Bir (test) programı geliştirilecektir. Testler, geliştirmenin ayrılmaz bir parçası olarak gerçekleştirildiğinden, genellikle ayrı bir test programı değil, test faaliyetlerini de içeren genel bir

program olacaktır. Program, örneğin sıralı yaşam döngülerinde olduğu gibi ayrıntılı dokümanite edilmiş bir program formatında olabilir, ancak örneğin Çevik yaşam döngülerinde olduğu gibi önceliklerin belirtildiği bir görev panosu formatında da olabilir.

Test planlamasının bir parçası olarak, çevrimdışı ve/veya çevrimiçi testleri gerçekleştirmek için gerekli bilgi ve becerilere sahip test personelinin mevcudiyeti için bir plan oluşturulur. Elbette bu, bu tür testleri gerçekleştirmek için gerekli bilgi ve becerilere sahip test uzmanları genellikle kıt bir kaynak olduğu için zor olabilir. Test personelinin test işini yapabilmesini sağlamak için test planının bir parçası olarak gerekli eğitimi tanımlamak gerekebilir.

2.2.5 SG5 Test Planına Bağlılık Sağlanması

Son olarak, test planı ve tanımlanan test yaklaşımı paydaşlar tarafından kabul edilmelidir. Paydaşlar, öncelikli ürün riskleri listesini ve gerçekleştirilecek test azaltma eylemlerini anlamalı ve kabul etmelidir. Onların anlayış ve taahhütleri, örneğin ürün risklerini, test yaklaşımını ve gerekçesini ve gerekli bütçeyi açıklayan kısa bir sunum ve ardından bir tartışma yoluyla sağlanabilir. Test planı ve test yaklaşımı ile ilgili tartışma, geliştirilecek MÖ modeli/Yapay zekâ sisteminin amacı bağlamında gerçekleştirilmeli ve iş öncelikleri ile uyum sağlanmalıdır. Kabul kriterleri de elbette paydaşlarla tartışılması ve üzerinde anlaşmaya varılması gereken bir başka önemli husustur.

Test planını ve test yaklaşımını geliştirmek ve oluşturmak için daha çevik bir yaklaşım kullanılırsa, bu zaten bir ekip çalışmasıdır ve muhtemelen bir yapay zekâ test uzmanı (ekip üyelerinden biri) tarafından yönetilir. Ürün kalitesi, ekibin sorumluluğundadır. Bu nedenle, ekip doğru süreci izlediği sürece, (test) yaklaşımına ve (test) planına bağlılık, bir ekip çalışması olduğu için zaten örtük bir sonuçtur. Elbette, ekip dışındaki diğer paydaşların da test planına bağlılık göstermesi gerekebilir; onlar için, daha önce de belirtildiği gibi, sunum ve tartışma oturumu önemli olmaya devam edecek ve gerçekleştirilmesi gerekecektir.

2.3 Süreç Alanı 2.3 Test İzleme ve Kontrol (TMC)

Test İzleme ve Kontrolünün amacı, test sürecinin ilerleyişi ve ürün kalitesi hakkında bilgi sağlamak ve böylece test sürecinin planından önemli ölçüde sapması veya ürün kalitesinin beklentilerden önemli ölçüde sapması durumunda uygun düzeltici önlemlerin alınabilmesini sağlamaktır.

2.3.1 SG1 Test İlerlemesinin Plan ile Karşılaştırılarak İzlenmesi

Test sürecini izlemek, ilerleme değerlendirmeleri yapmak ve test sürecine ilişkin raporlar hazırlamak, yapay zekâ tabanlı bir sistemi test etmek için geleneksel (yapay zekâ olmayan) bir sistemi test etmekten farklı değildir. Test sürecinin izlenmesinin uygulanma ve gerçekleştirilme şekli, büyük ölçüde uygulanan yaşam döngüsüne bağlıdır. Sıralı V modeli türündeki yaşam döngülerinde, test sürecinin izlenmesi genellikle daha resmi ve dokümantasyon yoğun bir süreçtir. Çevik türündeki yaşam döngülerinde ise test sürecinin izlenmesi genellikle çok daha az resmi olup, görev panosu ile desteklenir ve dokümantasyon gereksinimi azdır, bkz. [TMMiAgile].

2.3.2 SG2 Ürün Kalitesinin Plan ve Beklentilere Göre İzlenmesi

Ürün kalitesini izlerken, yapay zekâ tabanlı bir sistemi test etmenin geleneksel bir sistemi test etmekten farklı olduğunu anlamak önemlidir. Bu fark, ürün kalitesinin hangi yönlerinin izlenmesi gerektiği konusunda da etkilidir. Geleneksel sistemlerde “test”, hataları arama, testleri yürütme ve performansı ölçme gibi çeşitli süreçleri ifade eder. Yapay Zekâ’da ise, bu terim özellikle özel bir veri setine göre sistem performansının istatistiksel olarak değerlendirilmesini ifade eder. Temel olarak, izlenen ve raporlanan çıkış kriterlerinin türü doğası gereği farklıdır (bazı örnekler için 2.2.2 numaralı paragrafa bakınız). Özetle, ürün kalitesinin izlenmesi özellikle belirlenen ürün riskleri ve tanımlanmış çıkış kriterleri için gerçekleştirilir. Hataların izlenmesi, sistemlerin yapay zekâ içermeyen kısımları için gerçekleştirilebilir. Ürün kalitesinin izlenmesinin nasıl uygulanabileceği konusunda önceki paragrafa bakınız.

Yapay zekâ sistemlerinin ürün kalitesini izlemede bir diğer fark, sistemin performansını sürekli olarak izlemektir. Yapay zekâ sistemlerinde, sistemin canlı kullanımındaki kalitesini sürekli olarak gözden geçirmek ve yeniden değerlendirmek gerekir. Dokümanda daha önce de belirtildiği gibi, yapay zekâ tabanlı sistemlerde model kayması, veri kayması ve kavram kayması, zaman içinde dikkate alınması ve izlenmesi gereken faktörlerdir. Sapma belirtilerini erken tespit etmek için model performans ölçütlerini (örneğin doğruluk) sürekli olarak izlemek gerekir. Ürün kalitesini piyasaya sürüldükten sonra da izlemeye devam etmenin önemli nedenleri arasında, düşmanca saldırıları önlemek, veri zehirlenmesi ve kendi kendine öğrenme evrim davranışını önlemek gibi faktörler yer alır.

2.3.3 SG3 Düzeltici Faaliyetlerin Kapanışa Kadar Yönetilmesi

Test sürecinin izlenmesi (yukarıya bakınız) gibi, düzeltici eylemlerin yönetimi de yapay zekâ tabanlı bir sistemi test etmek için geleneksel (yapay zekâ olmayan) bir sistemi test etmekten farklı değildir. Ancak, sorunların türü farklı olabilir. Yapay zekâ sistemlerini test ederken karşılaşılabilecek tipik sorunlar, girdi verilerinin kalitesi ile ilgilidir. Veri kalitesi sorunları, proje üzerinde önemli bir etkiye sahip olabilir ve/veya düzeltilmesi uzun zaman alabilir.

Düzeltilici eylemlerin kapatılana kadar yönetilme ve gerçekleştirilme şekli, büyük ölçüde uygulanan yaşam döngüsüne bağlıdır. Sıralı V modeli türündeki yaşam döngülerinde, ilerlemeyle ilgili sorunlar genellikle bir test yöneticisi veya proje yöneticisi tarafından yönetilir. Çevik türdeki yaşam döngülerinde ise, ilerleme engellerinin üstesinden gelmek genellikle bir Scrum master’ın sorumluluğundadır. Bu tür düzeltici eylemlerin durumu bir görev panosunda gösterilir ve günlük standup toplantısında tartışılır.

2.4 Süreç Alanı 2.4 Test Tasarımı ve Yürütme (TDE)

Test Tasarımı ve Yürütme'nin amacı, test tasarım spesifikasyonunu oluşturmak, test tasarım tekniklerini kullanmak, yapılandırılmış bir test yürütme süreci gerçekleştirmek ve test olaylarını kapanışa kadar yönetmek suretiyle test tasarım ve yürütme sürecinde test süreci yeteneğini iyileştirmektir.

Amaç beyanı yüksek düzeyde hala geçerli olsa da, test koşullarını, test durumlarını tanımlama ve bunları yürütme süreci, aşağıda açıklandığı gibi çok farklıdır. Bir model oluşturulduktan sonra (yani, eğitildikten, değerlendirildikten ve ayarlandıktan sonra), kararlaştırılan fonksiyonel performans kriterlerinin karşılandığından emin olmak için bağımsız bir test veri seti ile test edilmelidir (bkz. bölüm 2.2.2).

2.4.1 SG1 Test Tasarım Tekniklerini Kullanarak Test Analizi ve Tasarımı Gerçekleştirme

Bir yapay zekâ sistemini test ederken, “test koşulları” ve “test tasarımı” kavramları geleneksel yazılım testlerinden temel olarak farklılık gösterir. Açık giriş-çıkış çiftleri üretmek yerine, odak noktası,

eđitilmiş modelin davranışını, doğruluđunu ve sađlamlıđını anlamlı bir řekilde deđerlendirebilecek veri kümeleri oluřturmak ve dođrulamaktır.

Birincil hedef, yapay zekâ modelinin çalıřacađı kořulları temsil eden veri kümelerinin tanımlanması ve tasarlanmasıdır. Bu, eđitim, dođrulama ve test için kullanılacak verilerin türlerini, kaynaklarını ve kalitesini belirlemeyi içerir. Tipik olarak, ortak bir veri kaynađından üç farklı ancak istatistiksel olarak temsil edici veri kümesi oluřturulur:

- Eđitim veri kümesi – YZ/MÖ modelini eđitmek için kullanılır.
- Dođrulama veri kümesi – model parametrelerini deđerlendirmek ve ayarlamak için kullanılır.
- Test veri kümesi (holdout veri kümesi) – ayarlanan modelin genelleme performansını deđerlendirmek için kullanılır.

TMMi bađlamında, test analizi ve tasarım ařaması, yapay zekâ sisteminin amacı, gereksinimleri ve risk deđerlendirmesinden elde edilen test verisi kořullarının ve deđerlendirme ölçütlerinin belirlenmesi ve önceliklendirilmesini içerir. Burada test temeli, yalnızca spesifikasyonları ve tasarım dokümanlarını deđil, aynı zamanda veri toplama stratejilerini, etiketleme kılavuzlarını ve model performans kriterlerini de içerir.

Uygun, tarafsız veriler genellikle sınırlı olduđundan, eđitim ve dođrulama veri kümeleri sıklıkla tek bir birleřtirilmiş veri kümesinden türetilirken, test veri kümesi izole kalmalıdır. Bu ayırım, objektif model deđerlendirmesini sađlar ve veri sızıntısını önler. Veri kümelerini bölmek için evrensel bir oran yoktur, ancak tipik bir kılavuz, eđitim, dođrulama ve test için 60:20:20'dir. Özel tabakalařma veya temsil kısıtlamaları (ör. küçük veri kümeleri veya kritik alt popülasyonlar) bulunmadıkça, veri bölünmesi rastgele olmalıdır. YZ'da test tasarımı, gerçeđçi ve uç durum senaryolarını yansıtan test durumlarının ve kořullarının sečilmesine ve önceliklendirilmesine odaklanır. Bu, ařađıdakileri içerebilir:

- Modelin zayıflıklarını arařtırmak için rakip veya stres testi verileri tasarlamak.
- Veri alanında eřdeđerlik bölümlendirme ve sınır testi kullanma (örneğin, özellik dađılımlarının kenar deđerleri).
- Çiftli test gibi kombinatoriyal test teknikleri kullanmak, çünkü yapay zekâ tabanlı bir sistem için ilgi çekici parametrelerin sayısı, özellikle sistem büyük veri kullandıđında veya dıř dünya ile etkileřime girdiđinde, son derece yüksek olabilir.
- Metamorfik test, başarılı olan bir kaynak test vakasına dayalı test vakaları oluřturmayı amaçlayan bir teknik olup, yapay zekâ sistemlerinin test edilmesinde kullanılabilir bir başka tekniktir.
- Test kořulları setinin bir parçası olarak yanlılık, tarafsızlık ve açıklanabilirlik testleri uygulamak.
- Veri kaynaklarından ve model gereksinimlerinden test veri kümelerine ve performans sonuçlarına kadar izlenebilirlik sađlamak.

Bu hedefin çıktıları arasında veri seti spesifikasyonu, test verilerinin kapsamının dokümente edilmesi ve tanımlanmış kabul kriterleri (ör. dođruluk eřikleri, güven aralıkları, yanlılık sınırları) yer alır. Bu unsurlar topluca, yapay zekâ modelinin test edilmesinin yapılandırılmış bir yaklařım izlemesini sađlar.

2.4.2 SG2 Test Uygulamasını Gerçekleřtirme

Yapay zekâ sistemleri için test uygulaması, test tasarımı sırasında tanımlanan test verilerinin ve ortamlarının elde edilmesi, hazırlanması ve sonlandırılmasını içerir. Yapay Zekâ testi, eđitilmiş modelin performansını, sađlamlıđını ve gereksinimlere uygunluđunu dođrulamak için kullanılacak veri kümelerini elde etmeye ve düzenlemeye odaklanır. Elde edilen veriler çeřitli biçimlerde olabilir (ör. sayısal, kategorik, görüntü, tablo, metin, zaman serisi, sensör, cođrafik, video veya ses) ve genellikle

gerçek testler için kullanılmadan önce kapsamlı bir ön işleme tabi tutulması gerekir. Ana ön işleme faaliyetleri arasında temizleme, dönüştürme, artırma ve örnekleme yer alır.

- Temizleme: Yanlış, yinelenen veya tutarsız veriler tespit edilir ve düzeltilir veya kaldırılır. Eksik veriler, tamamlama teknikleri kullanılarak ele alınabilirken, kişisel veya hassas bilgileri korumak için anonimleştirme veya takma ad kullanma yöntemleri uygulanır. Aykırı değerler, yapay zekâ sisteminin operasyonel bağlamıyla olan ilgilerine bağlı olarak kaldırılabilir veya saklanabilir.
- Dönüştürme: Veri formatları veya yapıları, test edilen modele uygun tutarlı bir biçime dönüştürülür (örneğin, kategorik verilerin sayısal biçime dönüştürülmesi, sayısal özelliklerin normalleştirilmesi veya görüntü formatlarının yeniden kodlanması). Bu dönüştürmeler, test uygulamaları arasında karşılaştırılabilirlik ve istikrarı sağlar.
- Artırma: Kapsam ve sağlamlığı iyileştirmek için, veri kümesi yeni veya değiştirilmiş örnekler (ör. döndürülmüş görüntüler, başka kelimelerle ifade edilmiş metinler veya gürültü eklenmiş veriler) oluşturularak yapay olarak genişletilebilir. Artırma, sistemin bozulmalara veya saldırılara karşı direncini test etmek için tasarlanmış karşıt örnekleri de içerebilir.
- Örnekleme: İstatistiksel geçerliliği korurken uygulanabilir testler sağlamak için büyük veri kümelerinden temsili alt kümeler seçilir. Örnekleme yöntemleri (rastgele, tabakalı veya risk temelli) çeşitliliği ve test tasarımında belirlenen kritik özellikleri korumalıdır.

Tüm ön işleme uygulama adımları dokümente edilmeli ve veri kümesi özelliklerine göre izlenebilir olmalıdır. Uygulanan test verileri, test komut dosyaları ve ortam yapılandırmaları, test yürütme ve model değerlendirmesi için gerekli olan test çalışma ürünlerini oluşturur.

2.4.3 SG3 Test Yürütme

Yapay zekâ sistemleri için test yürütme, hazırlanan test veri kümelerini ve komut dosyalarını çalıştırarak, eğitilmiş modelin tanımlanmış kabul kriterlerine göre nasıl performans gösterdiğine dair kanıt toplamayı içerir. Amaç, modelin amaçlandığı gibi davranmasını, beklenen performans düzeylerine ulaşmasını ve önceki aşamalarda belirlenen veri koşulları aralığında güvenilir bir şekilde çalışmasını sağlamaktır. Testler, yapay zekâ modelinin eğitimi, ayarlaması ve doğrulaması tamamlandıktan ve tanımlanmış kriterlere uygun olduktan sonra başlatılır. Bu, test yürütmeyi başlatmak için bir giriş kriteri olarak düşünülebilir.

Yapay zekâ sistemlerini test ederken, bir testi gerçekleştirmek genellikle test veri setini eğitilmiş ve doğrulanmış modele uygulamak ve çıktıları (tahminler, sınıflandırmalar, puanlar veya eylemler) kaydetmek anlamına gelir. Bu çıktılar, beklenen veya referans sonuçlarla (gerçek değer) karşılaştırılır ve karışıklık matrisleri ve doğruluk, kesinlik, geri çağırma veya F1 puanı gibi nicel performans ölçütleri kullanılarak değerlendirilir (bkz. paragraf 2.2.2).

Yapay zekâ sistemlerinin doğasında bulunan stokastik yapı nedeniyle, bireysel test uygulamaları biraz farklı sonuçlar verebilir. Bu nedenle, testler kontrollü koşullar altında tekrarlanmalı ve sonuçlar istatistiksel olarak analiz edilmelidir (örneğin, performans ölçütlerinin ortalaması ve varyansı). Bu, sonuçların tekrarlanabilir ve istikrarlı kanıtlara dayandığından emin olunmasını sağlar.

Tüm test sonuçları, anomaliler ve beklenen performanstan sapmalar dokümente edilmeli ve ilgili test verileri ve tanımlanan koşullara kadar izlenebilir olmalıdır. Test olayları, veri kalitesi sorunlarından, model sınırlamalarından veya uygulama hatalarından kaynaklanıp kaynaklanmadığını belirlemek için analiz edilir. Bu özel hedefin çıktıları arasında yürütülen test günlükleri, elde edilen sonuçlar, performans raporları ve olay raporları bulunur.

2.4.4 SG4 Test Olaylarını Kapanışa Kadar Yönetme

Test olaylarını kapanışa kadar yönetme süreci ve uygulaması, yapay zekâ sistemleriyle ilgili olaylar ve hatalar için de farklı değildir. Elbette daha ayrıntılı teknik düzeyde, bulunan hataların türü ve hata analizi gibi görevlerin yerine getirilme şekli vb. konularda farklılıklar vardır. Örneğin, ayrıntılı bir veri kalitesi analizi, doğruluğun azalması, yanlışlık bulunan model ve tehlikeye atılmış model ile ilgili hataları ortaya çıkarabilir.

Olay yönetimi süreci de büyük ölçüde uygulanan geliştirme yaşam döngüsüne bağlıdır. Geleneksel bir projede olayların dokümente edilmesi ve kapatılması ile Çevik bağlamındaki bir projede olayların dokümente edilmesi ve kapatılması arasında önemli farklılıklar vardır [TMMiAgile].

2.5 Süreç Alanı 2.5 Test Ortamı (TEV)

Test Ortamının amacı, testlerin yönetilebilir ve tekrarlanabilir bir şekilde yürütülmesinin mümkün olduğu uygun bir ortam oluşturmak ve sürdürmektir.

Test Ortamı süreç alanı ile, testlerin yönetilebilir ve tekrarlanabilir bir şekilde yürütülmesinin mümkün olduğu, genel test verilerini de içeren uygun bir test ortamı oluşturulur ve sürdürülür. Yapay zekâ tabanlı sistemler çok çeşitli operasyonel ortamlarda kullanılabilir, bu da test ortamlarının da benzer şekilde çeşitlilik gösterdiğini anlamına gelir. Test ortamlarının geleneksel sistemlerden farklı olmasına neden olan yapay zekâ tabanlı sistemlerin özelliklerine örnek olarak kendi kendine öğrenme, otonomi, çoklu kurum ve büyük veri verilebilir. Ayrıca, yapay zekâ tabanlı bir sistemi test ederken genellikle sanal test ortamları kullanılır. Bu, genellikle tehlikeli, olağandışı, aşırı ve zaman alıcı senaryoları test etmeyi kolaylaştırır. Sanal test ortamları ayrıca test ortamının çok daha fazla kontrol edilebilirliğini sağlar. Sanal test ortamları tarafından donanımın simülasyonu, sistemlerin başka türlü kullanılamayacak (simüle edilmiş) donanım bileşenleriyle test edilmesini sağlar.

Yapay Zekâ/MÖ sistem iş akışının genellikle büyük bir bölümünü oluşturan veri hazırlama işleminin, test tasarımı ve yürütme süreci alanının bir parçası olduğu unutulmamalıdır (bkz. paragraf 2.4). Veriler prensip olarak, ayarlanan modeli eğitmek, doğrulamak ve test etmek için kullanılacak (test) vakalarını kapsar. Bu nedenle, (test) verileri geleneksel sistemlere kıyasla yapay zekâ tabanlı sistemleri test etmek için çok önemli bir rol oynar.

2.5.1 SG1 Test Ortamı Gereksinimlerini Geliştirme

Yapay zekâ tabanlı bir sistem için test ortamı gereksinimlerinin belirlenmesi, projenin erken aşamalarında gerçekleştirilmelidir. Test ortamı gereksinimlerini belirleyenlerin, geliştirilecek yapay zekâ sistemini ayrıntılı bir şekilde anlaması gerekir, çünkü sistemin özellikleri (bazı örnekler için yukarıya bakınız) neyin gerekli olduğunu belirleyecektir. Gereksinimlerin belirlenmesi, doğruluğu, uygunluğu, uygulanabilirliği ve “gerçek hayattaki” çalışma ortamını doğru bir şekilde yansıtmayı yansıtmadığının kontrol edilmesi için gözden geçirilir. Erken gereksinimlerin belirlenmesi, gerekli test ortamını ve bileşenlerini edinmek ve/veya geliştirmek için daha fazla zaman sağlama avantajı sunar. Test ortamı genellikle çok basit olmayıp oldukça karmaşık olabileceğinden, bu özellikle yapay zekâ tabanlı sistemler için çok önemlidir.

2.5.2 SG2 Test Ortamı Uygulamasını Gerçekleştirme

Bu TMMi özel hedefi ile test ortamı gereksinimleri uygulanır ve test ortamı yapay zekâ modelini test etmek için kullanıma sunulur. Test ortamı bileşenleri üzerinde uygun şekilde birim testleri gerçekleştirilir ve uygulama aşamasının sonunda, test ortamının yapay zekâ modelini test etmek için

kullanıma hazır olup olmadığını belirlemek için bir güvenilirlik testi yapılır. Ortamı test ederken genel test verileri kullanılır.

2.5.3 SG3 Test Ortamlarını Yönetme ve Kontrol Etme

Test ortamının yönetimi ve kontrolü, bir yapay zekâ projesi boyunca gerçekleştirilecektir. Genellikle proje ekibi test ortamını kendisi kurmalı ve sürdürmelidir, ancak bu, ekip dışındaki ayrı bir birim tarafından da yapılabilir. Faaliyetler ekip tarafından gerçekleştirildiğinde, bu elbette ekibin bu görevleri yerine getirebilecek yeterli teknik bilgiye sahip olduğu anlamına gelir. Aynı zamanda, bu görevlerin projenin bir parçası haline geldiği anlamına da gelir; örneğin, planlama oturumunda ele alınmaları, görev panosuna yazılmaları ve herhangi bir sorun olması durumunda günlük ayakta (stand-up) toplantılarında tartışılmaları gerekir. Test Ortamı süreç alanı için ana sonuç olarak, belirli hedefler ve uygulamaların tümü geçerlidir ve özünde bir değişiklik yoktur. Farklılık, test ortamının türü, test verileri ve gerekli bileşenlerde olabilir.

Referanslar

[AiU] Artificial Intelligence United (2019), AiU Sertifikalı Yapay Zekâ Test Uzmanı (CTYZ) Müfredatı V1.0, Artificial Intelligence United

[ISTQB] Uluslararası Yazılım Test Yeterlilik Kurulu (2021), Sertifikalı Test Uzmanı Yapay Zekâ Testi (CT-YZ) Müfredatı V1.0, ISTQB

[Smith] A.L. Smith, R. Black, J. Davenport, J. Olszeweska, J. Röβler ve J. Wright (2022), Yapay Zekâ ve Yazılım Testleri, BCS Öğrenme ve Gelişim,

[TMMiAgile] E. van Veenendaal (ed.) (2020), Çevik Bir Dünyada TMMi V1.4, TMMi Vakfı