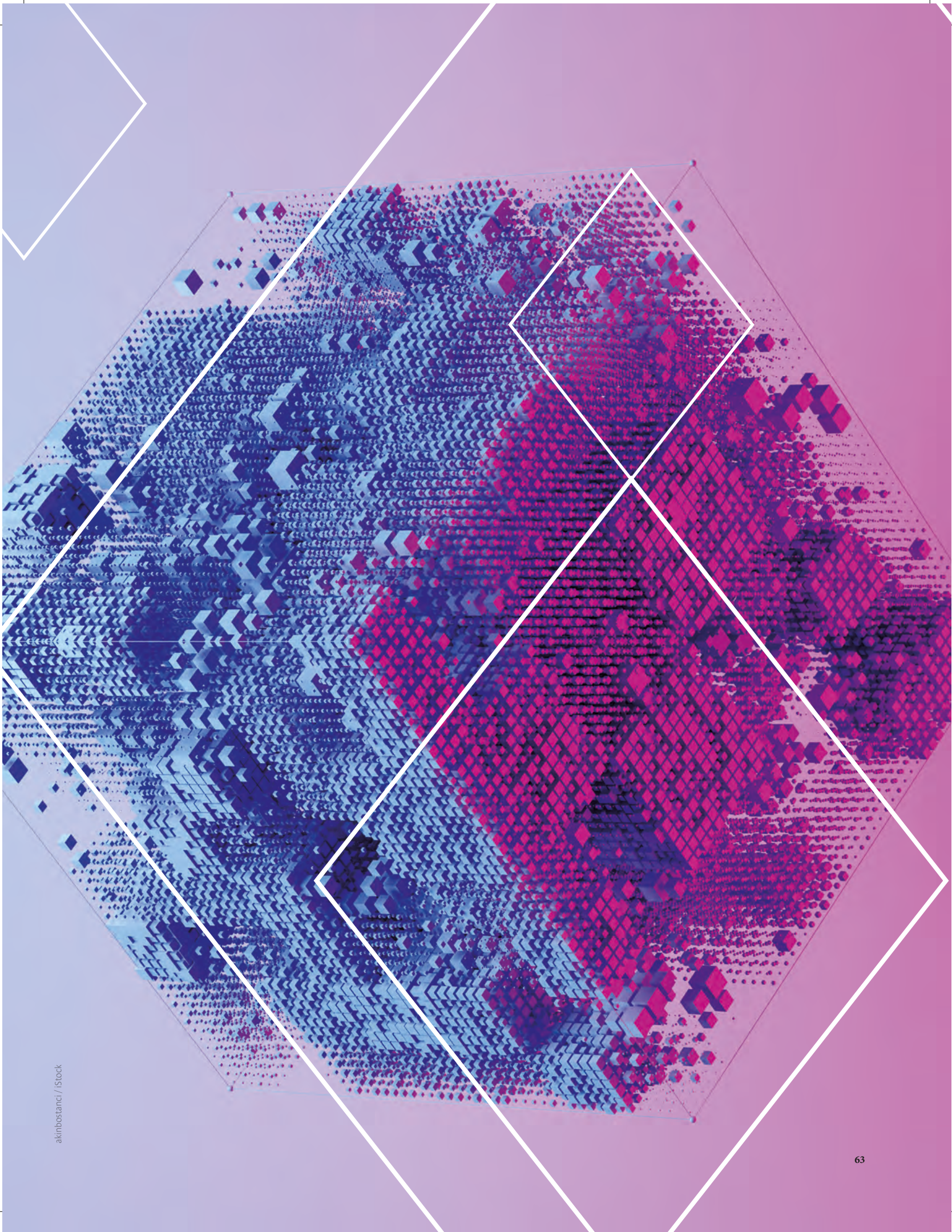


Yapay Zekâda Akıl Yürütmenin Sınavı

Murat Can Işık [*Bilgisayar Mühendisi, Makine Öğrenmesi Donanımı, Stanford Üniversitesi*]

Yapay zekâ artık yalnızca teknoloji şirketlerinin laboratuvarlarında değil, gündelik hayatımızın her alanında etkili bir şekilde kullanılıyor. Telefonun sabah alarmı ertelerken ekran parlaklığını otomatik ayarlaması, kameranızın fotoğrafı çektikten sonra saniyeler içinde görüntüyü netleştirmesi, navigasyonun trafik durumuna göre rotayı değiştirmesi... Bunların hepsi aslında arka planda çalışan yapay zekâ benzeri sistemlerin eseridir.

Bu gelişmelerin ardından ortaya çıkan en önemli tartışmalardan biri şudur: Yapay zekâ gerçekten akıl yürütebiliyor mu yoksa sadece daha önce öğrendiği örnekleri mi tekrar ediyor?



akimbostanci / iStock

Yapay Zekânın Akıl Yürütme Becerisi Nasıl Ölçülür?

İnsan zekâsı yalnızca bilgiyi depolama değil, onu yeni durumlara uyarlama becerisi olarak da tanımlanır. Bir insan daha önce hiç görmediği bir bulmacayı bile az sayıda örnekten kural çıkararak çözebilir.

Günümüzde yapay zekâ sistemleri özellikle de büyük dil modelleri (LLM) çok büyük miktarda veriyle eğitilerek etkileyici sonuçlar üretebiliyor. Fakat bu başarı çoğu zaman önceden öğrenilen bilgilerin hatırlanması temellidir. Model daha önce çok miktarda benzer örnekle eğitildiyse

başarılı olur ama yetersiz veriyle eğitildiğinde ya da yeni bir durumla karşılaştığında çoğu zaman başarısız kalır. İşte bu noktada ARC-AGI devreye girer.

2019'da François Chollet tarafından geliştirilen Yapay Genel Zekâ için Soyutlama ve Mantık Yürütme Testi (ARC-AGI) yapay zekânın “akışkan zekâ” denilen özelliğini ölçer. Bu kavram bir sistemin sınırlı deneyimle yeni bir problemi çözebilme becerisini tanımlar. Yani bir modeli binlerce örnekle eğitmek yerine modelin az sayıda örnekten kural çıkarma yani veriler arasındaki örüntüleri belirleme yeteneğini test eder. Bu sayede sınırlı sayıda örnekten yola çıkarak yeni durumlara uyum sağlayabilen sistemlerle yalnızca büyük veri kümeleri

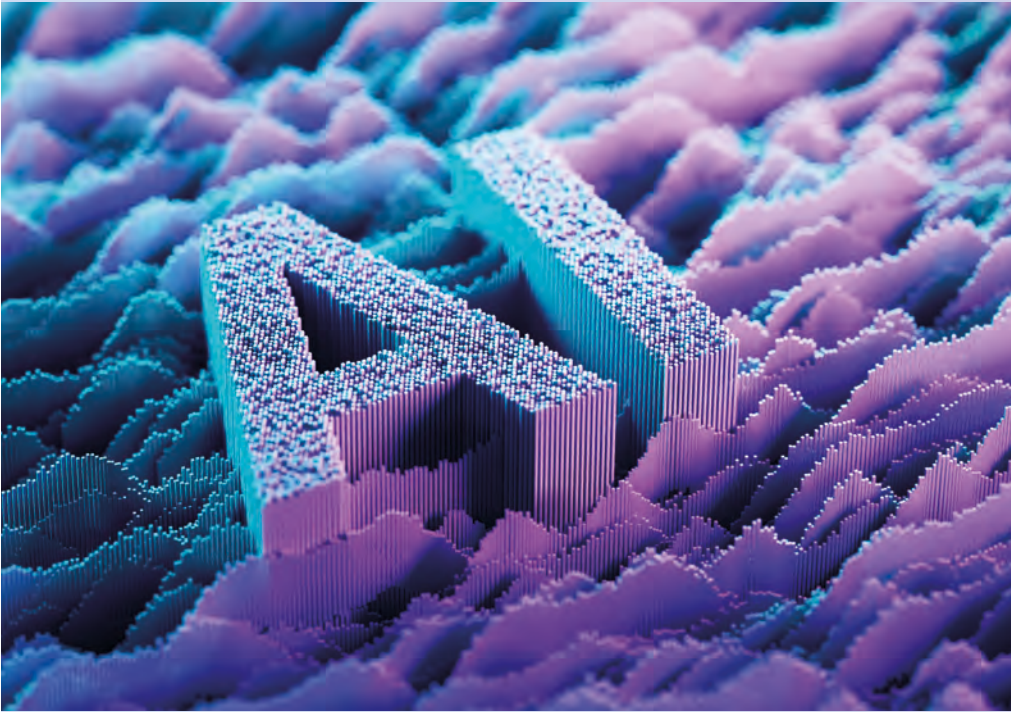
üzerinden bilgiyi hatırlayan sistemleri birbirinden ayırmak mümkün olur.

Tarihsel olarak bu yeni bir fikir değildir. 20. yüzyılın ortalarında insan zekâsını ölçmek için geliştirilen bazı psikometrik testler de (örneğin Raven Progresif Matris Testleri) benzer bir mantığa dayanıyordu. Bu testlerde katılımcılardan az sayıda örneği inceleyip eksik parçayı tahmin etmeleri istenirdi. ARC-AGI da bu yaklaşımı kullanarak yapay zekâ sistemlerini değerlendiren modern bir ölçektir. ARC-AGI, görünüşte etkileyici sonuçlar üreten uygulamalarının ötesinde yapay zekânın gerçek akıl yürütme becerisini sınamak için önemli bir araç hâline geldi.

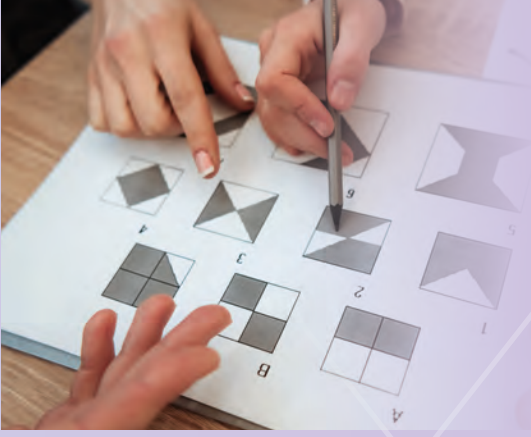
Görevler Nasıl Çalışır?

ARC-AGI testindeki görevler bir tür “mantık bulmacası” gibidir. Her görevde yalnızca 2 ya da 3 girdi ve çıktı çifti örnek olarak verilir. Modelin görevi bu örnekler arasındaki örüntüyü çözerek verilen yeni bir girdiye doğru çıktıyı üretmektir.

Örneğin küçük renkli karelerden oluşan uzgaralar üzerinde tanımlan bir



mustafa / iStock



T. Turovska / iStock

görevde mavi karelerin kırmızı bir çizgiye göre ayna simetrisi alınmış olabilir. İnsan beyni bu örüntüyü fark eder ve yeni bir durumda aynı kuralı uygular. Bir başka görevde belirli renkteki karelerin birleştirilip yeni bir şekil oluşturulması gerekebilir. İnsanlar hatta çocuklar bile bu tür örüntüleri az sayıda örnekten yola çıkarak hızlıca fark edebilir yani sezgisel olarak kavrayabilir.

Bir yapay zekâ modeli içinse durum çok daha farklıdır. Çünkü önceden bu kuralı öğrenmemiştir. Elindeki tek bilgi, örneklerde gördüğü girdi-çıkı çiftleridir. Bu kısıtlı bilgilerle doğru kuralı bulmak güçlü bir problem çözme becerisi gerektirir.

ARC-AGI'nın önemli bir özelliği testte kullanılan görevlerin internette kolayca bulunamamasıdır. Dolayısıyla akıl yürütme becerisi ölçülecek bir model bu görevleri daha önce öğrenmiş olamaz. Böylece ezberleme yani daha önce öğrendiği bilgilerden hatırlama

olasılığı ortadan kaldırılmış olur. Modelin kuralı örneklerden çıkarması gerekir.

Bu yaklaşım aslında insan zekâsının temel özelliklerinden biri olan “transfer” yeteneğini test eder. Transfer, öğrenilen bir bilgiyi yeni bir durumda kullanabilmektir. Bir öğrenci matematikte bir formülü öğrendikten sonra onu başka bir soruya uygulayabiliyorsa transfer yeteneğine sahip demektir. ARC-AGI da yapay zekâ sistemlerinin bu kabiliyetini ölçer.

Neden Zor?

ARC-AGI testleri ilk bakışta basit görünebilir. Ancak bu görevler yapay zekâ modelleri için ciddi birer sınavdır. Bunun üç ana nedeni vardır:

Birinci neden:

Önceki bilgileri hatırlamak işe yaramaz. Modern yapay zekâ sistemlerinin çoğu büyük veri kümelerine dayanır. Milyonlarca örnekle eğitilen bir sistem daha önce öğrendiğine benzeyen görevlerde çok başarılı olabilir. Ancak ARC-AGI az örnekle çalışır ve modelin daha önce bu görevle ilgili bir bilgisi yoktur.

İkinci neden:

Kural uzayı yani olası kuralların oluşturduğu küme çok geniştir. Bir görevi çözmek için yansıtma, döndürme, renk değiştirme,

sayma, birleştirme veya silme gibi yüzlerce farklı olasılığın denenmesi gerekebilir.

Modelin her görevde doğru kuralı bulabilmesi için kural uzayındaki denenebilecek kombinasyonların sayısı hızla artar. Bu durum “arama uzayı patlaması” denen bir soruna yol açar. Bir insan sezgileriyle bu arama alanını daraltabilir ama bir yapay zekâ modeli bunu yapamazsa sonuçlara ulaşması çok uzun sürebilir.

Üçüncü neden:

İnsanlar şekilleri, simetriyi, renkleri, mekânsal ilişkileri doğal olarak kavrar.

Bu durum yapay zekâ sistemleri için geçerli değildir. Her kavram açıkça tanımlanmalıdır ve modelin verilen örnekler üzerinden aralarındaki örüntüleri keşfetmesi gereklidir.

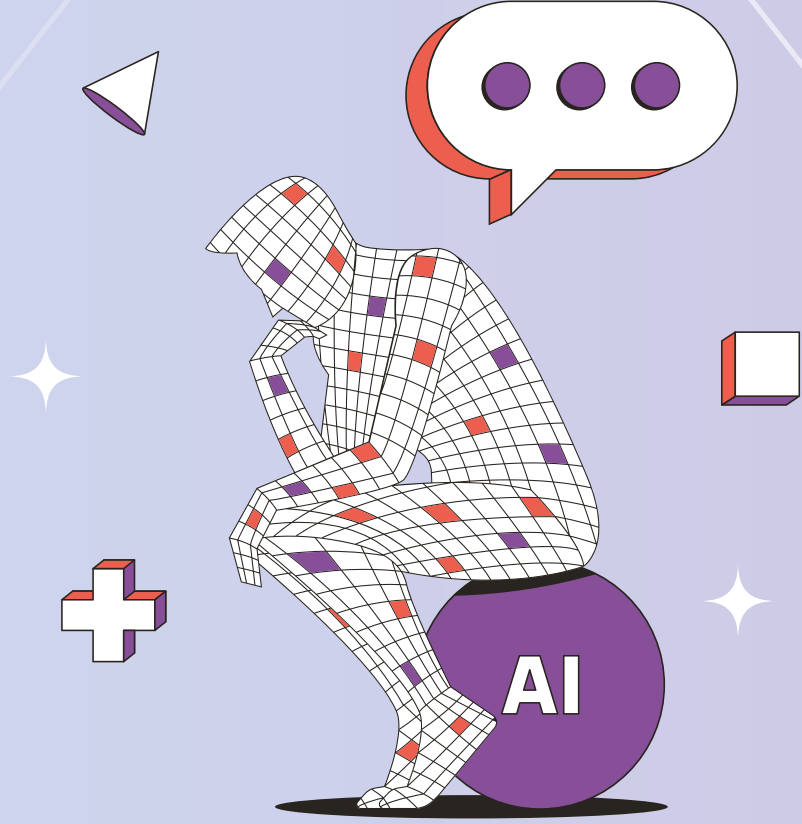
Bu yüzden ARC-AGI testlerinde şu anki büyük yapay zekâ modellerinin başarı oranları oldukça düşüktür. Örneğin insanlar bu görevlerin büyük kısmını birkaç denemede çözebilirken yapay zekâ sistemleri hâlâ %10-15 civarında başarı gösteriyor. Bu durum, yapay zekânın insan zekâsının becerilerine yaklaşmak için ne kadar yol alması gerektiğini açıkça gösteriyor.

Etkili Yaklaşımlar ve Güncel Durum

ARC-AGI yarışmalarında başarılı olan yaklaşımlar, önceki bilgileri hatırlamaktan ziyade stratejiye dayanır. Bu stratejilerden bazıları şunlardır:

- Görsel özellikleri ayrıştırma: Renk, şekil, boyut ve konum gibi özellikleri ayrı ayrı işlemek.
- Kural adayları üretme: Küçük bir programlama dili (DSL) kullanarak olası kuralları sistemli biçimde oluşturmak.
- Arama + sezgi hibriti: Bazı modeller daha önceki görevlerde öğrenilmiş örüntüleri kullanarak daha az kural dener ve doğru çözüme daha hızlı yaklaşır.
- Çok modüllü yapılar: Görsel analiz, semboller üzerinden akıl yürütme ve denetleme gibi farklı bileşenlerin birlikte çalıştığı modüller geliştirilir.

Özellikle son yıllarda bu stratejileri kullanan büyük modeller (örneğin Grok-4) ARC-AGI testlerindeki görevlerde daha iyi sonuçlar almaya başladı. Bu modeller yalnızca bir cevabı tahmin etmekle kalmayıp bir dizi mantıksal işlem zinciri kurabiliyor. Bu gelişmelere



rağmen yapay zekâ sistemlerinin insan zekâsının düzeyine ulaşması için katetmesi gereken ciddi bir mesafe var.

ARC-AGI testlerindeki görevleri çözebilen yapay zekâ sistemlerinin geliştirilmesini desteklemek amacıyla ARC Prize adlı bir yarışma gerçekleştiriliyor. 2024 ARC Prize yarışmasında, verilen görevlerin çözümünde kullandıkları kodları açık kaynak olarak paylaşan takımlardan en başarılı olanları bu görevlerin yaklaşık %50'sini doğru şekilde çözebildi. İnsanların aynı görevlerdeki başarı ortalaması ise %85'in üzerindeydi. Bu fark kural çıkarma, bağlamı

anlama, genelleme yapma gibi insan zekâsına özgü becerilerin yalnızca büyük miktarda veriyle eğitilmekle kazanılamayacağını gösteriyor.

Bir diğer önemli nokta da verimlilik ölçütüdür. Bir model bir kuralı 100 denemede buluyorsa bu verimli bir yapay zekâ örneği olmadığını gösterir. ARC-AGI'da hedef en az denemeye doğru kuralı bulabilmektir. Bu, insan zekâsına daha yakın bir ölçüm sağlar.

Türkiye İçin Fırsatlar

ARC-AGI yalnızca bilim insanları için değil, eğitim, sanayi ve araştırma alanlarında da büyük fırsatlar sunar.

Öğrenciler için basit ARC-AGI etkinlikleri tasarlanabilir. Örneğin 5x5 ölçülerindeki karelerden oluşan ızgara yapıları üzerinden tasarlanan basit görevlerde öğrenciler kural keşfi yapabilir. Bu etkinlikler matematiksel düşünme, örüntüleri fark etme ve problem çözme yeteneklerini güçlendirir. Ayrıca algoritmik düşünme (belirli bir problemi çözmek veya belirli bir amaca ulaşmak için çözüm yolunun adım adım tasarlanması) becerisini erken yaşta kazandırmak için etkili bir araçtır.

Üretim süreçlerinde otomasyon, robotik sistemler, sürücüsüz araçlar gibi alanlarda yapay zekâ sistemlerinin yeni durumlara hızlı uyum sağlaması gerekir. ARC-AGI yaklaşımına dayalı olarak geliştirilecek modüller, sistemlerin yeni durumlara uyum sağlamasını ve kuralları kendi kendine keşfedebilmesini sağlar. Bu da üretimde esneklik ve hız kazandırır.

Ülkemizde ARC-AGI testlerindeki görevleri çözebilen açık kaynaklı yapay zekâ sistemleri ve Türkçe yazılım araçları geliştirilebilir.

dem10 / iStock

Üniversite-sanayi iş birliğiyle geliştirilen yerli yapay zekâ sistemleri ile hem ulusal hem uluslararası yarışmalara katılım sağlanabilir. Bu sayede Türkiye, yapay zekâ alanındaki yükselen eğilimleri zamanında fark ederek bu alanda yeni yaklaşımların şekillenmesine katkı sunan öncü ülkelerden biri olabilir.

Sonuç olarak zekâyı yalnızca bilgi birikimiyle değil kural çıkarma gücüyle değerlendirmek gerekir. ARC-AGI, bu ihtiyaca yönelik geliştirilen en sistematik

çözümlerden biridir. Bu ölçüt, yapay zekâ teknolojilerinde insan zekâsının kapasitesine ulaşma konusunda hangi aşamada olduğumuzun anlaşılmasına yardımcı olur.

Yapay zekânın geleceğini şekillendiren en önemli unsurlardan biri artık veri miktarı değil, az veriden çok şey öğrenme becerisi olacaktır. Bu nedenle ARC-AGI gibi ölçekler önümüzdeki yıllarda daha da önem kazanacaktır. ■

Kaynaklar

François Chollet, "On the Measure of Intelligence" (arXiv) <https://arxiv.org/abs/1911.01547>
ARC-AGI tanıtım ve belgeler (ARC Prize) <https://arcprize.org/>
ARC-AGI görev aileleri ve örnekler (ARC Prize) <https://arcprize.org/arc>
ARC GitHub deposu, veri seti ve yönergeler (GitHub) <https://github.com/fchollet/ARC>
ARC Prize 2024-2025 sonuçları ve teknik rapor özetleri (ARC Prize) <https://arcprize.org/>
Verimlilik ve başarı ilişkisini gösteren panolar (ARC Prize) <https://arcprize.org/leaderboard>
Araçla akıl yürütme vizyonu bağlamında güncel model duyuruları, örnek: Grok-4 (xAI) <https://x.ai/grok>