

## Makineler İnsanlara Karşı

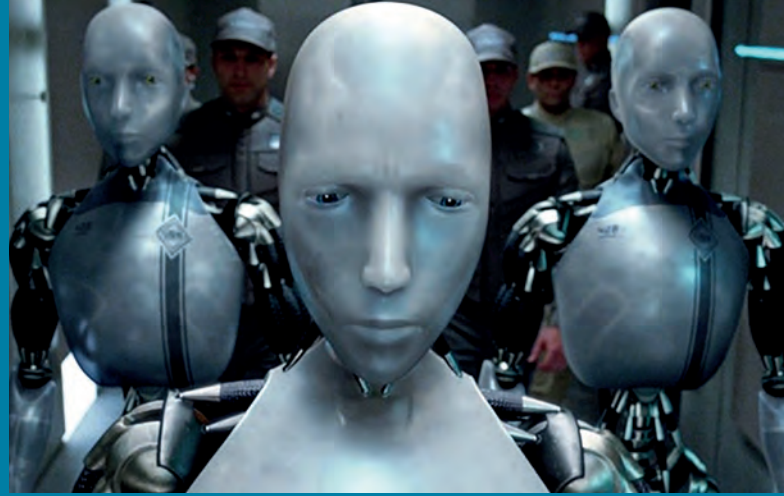
Yapay zekâ alanındaki yeni gelişmeler ünlü bilim kurgu yazarı Isaac Asimov'un eserlerinde geçen üç robot yasasını tekrar gündeme getirdi. Üç robot yasası şu şekilde yazılmıştı:

- Bir robot, bir insana zarar veremez ya da zarar görmesine seyirci kalmaz.
- Bir robot, birinci kuralla çelişmediği sürece bir insanın emirlerine uymak zorundadır.
- Bir robot, birinci ve ikinci kuralla çelişmediği sürece kendi varlığını korumakla mükelleftir.

Makinelerin gelişerek başlarına buyruk hareket etmesi ve insanlarla mücadeleye girişmeleri *Matrix*, *Terminatör* gibi bilim kurgu filmlerinde de çokça işlenmiş bir konu.

Yapay zekânın birçok alanda kullanılmaya başlanması, bir bilim kurgu ögesi olan insanlarla robotların savaşının o kadar da uzak bir ihtimal olmadığını gösteriyor. Bu çerçevede Google DeepMind ve Oxford İnsanlığın Geleceği Enstitüsü'nden araştırmacılar Laurent Orseau ve Stuart Armstrong robotların güvenli bir şekilde durdurulmasını sağlayacak bir altyapı geliştirdiklerini duyurdu.

32. Yapay Zekâda Belirsizlik Konferansı'nda yayımlanan makalede insanlarla birlikte çalışan robotların istenmeyen davranışlar sergilediklerinde durdurulmasını sağlayan bir mekanizma olması gerektiği belirtiliyor. Normal şartlarda "kırmızı bir düğmeye" basarak durdurulabilecek robot, uzun vadede öğrendikleri sayesinde bu düğmenin kendi çıkarına olmadığını "anlayıp" onu devre dışı bırakabilir ve durdurulamaz hale gelebilir. Bunu engellemek için yapay zekânın öğrenmesini yani insanlar tarafından durdurulduğu durumlardan ders çıkarmasını engelleyecek bir algoritma öneriliyor.



<https://www.fhi.ox.ac.uk/google-deepmind-and-fhi-collaborate-to-present-research-at-uaf-2016/>

Geçtiğimiz aylarda Elon Musk da robotların isyan edip insanlarla karşı karşıya gelmesinin önümüzdeki on yıl içinde gerçekleşebilecek bir durum olduğunu belirtmişti.

Bu tür öngörülerin ne kadar gerçekçi olduğu tartışılmalı da bilim insanlarının bu konuda çalışma başlatmış olması yapay zekânın ulaştığı noktayı göstermesi açısından dikkate değer.

