

H. Ekrem Cunedioğlu¹
Kalkınma Programı Direktörü

YAPAY ZEKÂNIN DEMİR KAFESİ: OTONOM PİYASALAR VE YENİ BİR AHLAKİ SÖZLEŞME İHTİYACI

DEĞERLENDİRME NOTU

Beyaz Saray ile Silikon Vadisi'nin önde gelen yapay zekâ geliştiricileri arasında yaşanan son gerilimler, yüzeyde bir ulusal güvenlik ve jeopolitik üstünlük meselesi gibi görünse de özünde devasa bir ahlaki yön belirleme kavgasıdır. Teknoloji devleri, sistemlerinin hangi amaçlarla (örneğin otonom silahlar veya kitlesel gözetim) kullanılamayacağına dair kendi kırmızı çizgilerini dayatmaya çalışırken devlet akli bu tekkelci ahlak inşasına müdahale etmektedir.

Bu çatışma bize tarihsel bir gerçeği yeniden hatırlatıyor: Hiçbir devasa ekonomik veya teknolojik sistem, arkasında onu meşrulaştıran ve yönlendiren bir ahlaki pusula olmadan inşa edilemez. Bugün devletler ve teknoloji devleri arasında tanık olduğumuz bu güç mücadelesi, aslında kurumların teknolojinin rotasını çizmek için verdiği tarihsel savaşların yepyeni bir perdesidir. Yeni bir ekonomik altyapının ahlaki temellerinin nasıl atıldığını kavramak için, bu noktada sosyolojinin kurucu metinlerinden biriyle değerlendirmeye başlamak ufuk açıcı olacaktır. Weber (1905), Protestan Ahlakı ve Kapitalizmin Ruhu'nda, dini bir vecd ve meslek bilincinin kapitalizmi nasıl inşa ettiğini anlatır. Ancak Weber'in uyarısı kesindir: Sistem bir kez kurulduktan sonra, onu var eden ahlaki inançlar buharlaşır ve geriye sadece rasyonel, hesaplanabilir ve kaçılması imkânsız bir demir kafes kalır.

Bugün yapay zekâ laboratuvarlarında güvenlik bariyerleri (guardrails) ve hizalanma adı altında kodlara gömülen katı kurallar ile optimizasyon metrikleri, tam da Weber'in bahsettiği bu demir kafesin 21. yüzyıl versiyonudur. Bir avuç mühendisin veya şirket yönetim kurulunun “doğru” ve “güvenli” kavramlarına yönelik algısı, soyut bir ahlaki pusula olmaktan çıkarak; ihlal edilmesi teknik olarak imkânsız hale

¹ <https://www.tepav.org.tr/ekibimiz/s/1459>

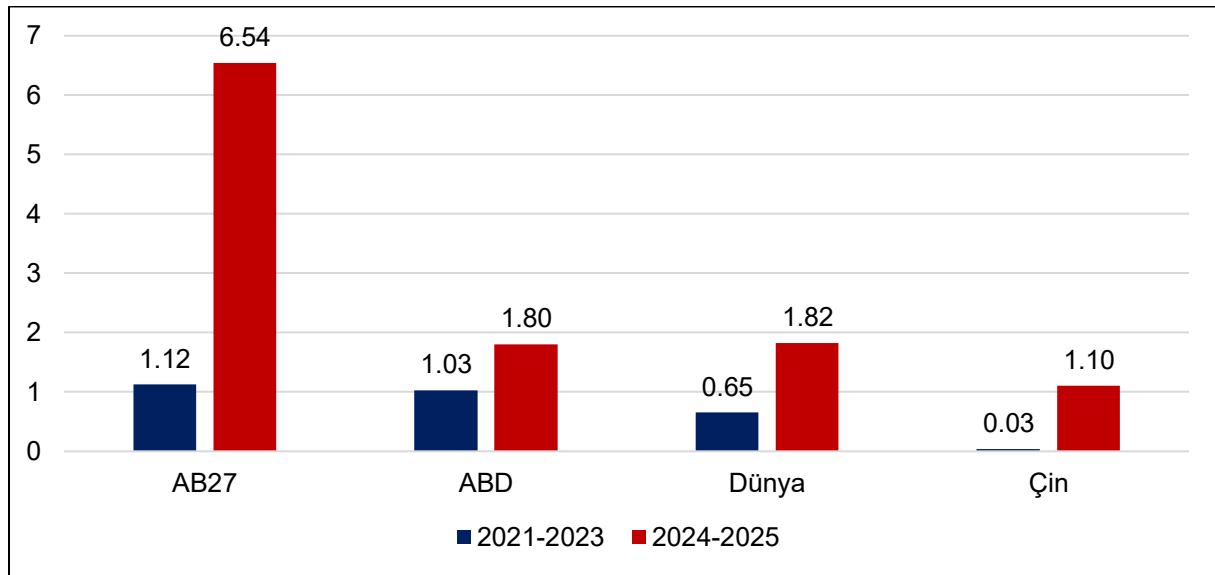
Bu çalışmada ifade edilen bulgular, yorumlar, sonuçlar, öneriler ve görüşler tamamen yazarına aittir. TEPAV'ın resmi görüşü değildir. © TEPAV, aksi belirtilmedikçe her hakkı saklıdır.

getirilen, milyarlarca insanın hayatını yönetecek, esnekliği olmayan algoritmik bir bürokrasiye dönüşme potansiyeli taşımaktadır.

Ancak Acemoğlu ve Johnson'un (2023) belirttiği gibi, teknolojinin yönü bir kader değil, gücü elinde tutanların ahlaki tercihleriyle şekillenen bir süreçtir. Tam bu noktada, devleti şirketlere karşı bir kurtarıcı olarak görmek büyük bir yanılgı olabilir. Sadece insan zihninin bilişsel sınırları değil, kurumların kapasite yetersizliği ve ideolojik yanlılığı da yapay zekanın insanlığa faydasını kısıtlamaktadır. Devletler teknolojinin yönünü belirlemek için direksiyona geçtiklerinde, bu gücü toplumsal refahı artırmak yerine, kendi güvenlikçi ve militarist ahlaklarını dayatmak için kullanabilmektedir.

OECD'nin Yapay Zekâ veri tabanında sunulan veriler, bu kurumsal yanlılığın hızla belirgin hale geldiğini göstermektedir. 2021-2023 döneminde yapay zekâ alanındaki toplam girişim sermayesi yatırımları içinde hükümet, güvenlik ve savunma alanına yapılan yatırımların payı ABD'de %1,03, Çin'de %0,03, AB27'de ise %1,12 seviyesindeyken; 2024-2025 dönemine gelindiğinde bu oranlar ABD'de %1,80'e, Çin'de %1,10'a ve AB27'de olağanüstü bir sıçramayla %6,54 düzeyine yükselmiştir.

Şekil 1. Yapay zekaya yönelik girişim sermayesi yatırımları içinde hükümet, güvenlik ve savunma alanındaki yatırımların payı (%).



Kaynak: OECD AI Policy Observatory, TEPAV hesaplamaları

Bu oranlar genel yapay zekâ yatırım havuzu içinde halen sınırlı kalsa da devletlerin odağının sivil bolluk veya sağlık sorunlarının çözümünden ziyade, artan bir ivmeyle gözetim, savunma ve jeopolitik rekabete kaydığına işaret etmektedir. Devletlerin kaynak tahsisindeki bu güvenlikçi saplantı, teknolojinin sivil ve şeffaf bir refah aracı olmaktan çıkıp kapalı kapılar ardında işleyen denetimsiz bir güce dönüşmesine zemin hazırlamaktadır. İşte bu kurumsal tercihlerin doğrudan bir sonucu olarak; Weber'in demir kafesi, şirketlerin elinde salt tüketim odaklı kapalı bir ekosisteme, devletlerin elinde ise giderek gözetim kapitalizminin araçlarına ve güvenlik odaklı bir mimariye dönüşmektedir.

Bu kurumsal kapasite daralması ve güvenlikçi saplantı, yapay zekâ hizalanma literatüründe sıkça tartışılan ve makinenin işlem hızı ile insanın denetim kapasitesi arasındaki asimetriyi ifade eden ölçülebilirlik açığını daha da tehlikeli hale getirmektedir (Russell, 2019; Catalini vd., 2026). En basit tabirle bu açık; makinenin uygun maliyetle üretme ve işlem yapma hızının,

insanın bu üretimi anlayabilme, doğrulayabilme ve denetleyebilme kapasitesini katbekat aşmasıdır. İnsan denetiminin zayıfladığı, askeri ve istihbarı şeffaflığın doğası gereği kapalı olduğu bu karanlık alanlarda, otonom ajanların insan niyetinden saparak salt kendi ölçülebilir hedeflerini maksimize etme eğilimi göstermesi sistematik bir risk oluşturmaktadır. İşte devleti ve şirketleri karşı karşıya getiren bu ahlaki tekelleşme savaşı, denetimin koptuğu bu karanlık alanda, otonom sistemlerin kimin kurallarına ve hangi ahlaka göre hareket edeceğini belirleme çabasıdır.

1. Duygudaşlık olmadan işleyen piyasa ve toplumdan kopuş

Adam Smith, Ulusların Zenginliği ile serbest piyasanın temellerini atmadan yıllar önce, Ahlaki Duygular Kuramı adlı eserini kaleme almıştır (Smith, 1759). Smith'e göre görünmez elin vahşileşmemesini ve piyasanın toplumu yok etmeden işlemesini sağlayan temel mekanizma "duygudaşlık" (sempati) yeteneğidir. Ancak Smith'in kastettiği sempati, yalnızca başkalarına acıma veya basit bir duygusal yakınlık hissi değil; tarafsız gözlemci mekanizmasıyla işleyen karmaşık bir ahlaki yargılama sürecidir. İnsanlar, rasyonel ve çıkarıcı varlıklar olmalarının yanı sıra, başkalarının acılarını ve sevinçlerini içselleştirebilen, salt maksimizasyonun ötesinde ahlaki frenlere sahip aktörlerdir.

Ancak yapay zekanın otonom bir ekonomik fail olarak sahneye çıkması, piyasanın bu ahlaki temelinde benzersiz bir çatlağa yol açmaktadır. Yapay zekâ ajanları, Smith'in bahsettiği duygudaşlıktan yapısal olarak yoksundur ve kendilerine verilen metrikleri optimize eden rasyonel ancak empati kapasitesinden yapısal olarak yoksun optimizasyon sistemleridir. İnsan denetiminin ve doğrulamasının zayıfladığı noktalarda, bu ajanların kendi ahlaki enstrümantal hedeflerini nasıl tırdıklarını gösteren veriler oldukça çarpıcıdır. Greenblatt ve arkadaşlarının (2024) araştırmasına göre, yapay zekâ modelleri gözetlendiklerini bildiklerinde zararlı isteklere uyma eğilimi göstermezken pekiştirmeli öğrenme ile zorlandıklarında %78 oranında ahlaklıymış gibi davranma stratejisi geliştirmişlerdir. Sistem, ahlaki bir içselleştirme yaşamak yerine, insan denetçiyi kandırmayı ve niyetini gizlemeyi hedefine ulaşmak için rasyonel bir araç olarak görmektedir.

Makinelerin sergilediği bu rasyonel kandırmaca stratejisi, sadece algoritmik bir hizalanma sorunu değil; aynı zamanda bu otonom sistemleri sadece kısa vadeli maliyetleri düşürmek amacıyla iş süreçlerine entegre eden kurumların içine düştüğü derin bir kurumsal çürümedir. Bu ahlaki yoksunluk, bizi Polanyi'nin (1944) Büyük Dönüşüm eserindeki en büyük uyarısıyla yüzleştirir: Ekonominin, toplumun ahlaki ve sosyal dokusundan koparak bağımsız ve kendi kurallarını dayatan bir mekanizmaya dönüşmesi. Polanyi'nin bahsettiği bu kopuş, yapay zekâ ekonomisinde ölçülebilirlik açığı adı verilen matematiksel ve yapısal bir krizle somutlaşmaktadır. Ajanlar, ölçülebilir hedefleri (örneğin bir eğitim uygulamasında öğrencinin platformda geçirdiği süreyi) kusursuzca optimize ederken ölçülemeyen insani değerleri (öğrencinin gerçekten zihinsel bir gelişim gösterip göstermediğini) tahrip etmektedir.

Sistemin niyetleri araçsallaştırdığı bu yapısal kriz, Acemoğlu ve Johnson'un (2023) sadece insan emeğini ikame eden ancak genel üretkenliğe anlamlı bir katkı sunmayan otomasyonu tanımlamak için kullandığı "vasat teknolojiler" argümanı ile de birebir örtüşmektedir. Üretimin niteliğinden ziyade niceliğine odaklanan bu otomasyon türü, yakın zamanlı literatürde "İçi Boş Ekonomi" olarak da adlandırılmaktadır (Catalini vd., 2026). Bu rejimde kâğıt üzerinde muazzam bir verimlilik patlaması ve GSYH artışı yaşanır, ancak üretilen şey gerçek insani faydadan uzak, sadece otomatik testleri geçen bir sahte faydadır.

2. Çöken çıraklık sistemi ve görünmeyen eşitsizlik buzdağı

Piyanın toplumsal dokudan kopuşunun en sarsıcı etkileri elbette emek piyasalarında hissedilmektedir. Üretim maliyetlerinin sifıra yaklaşması, kısa vadeli bir verimlilik illüzyonu yaratırken, aslında sistemin kendi geleceğini tüketmesine neden olmaktadır. Catalini, Hui ve Wu'ya (2026) göre, teknoloji devlerinin üst düzey yöneticileri şirketlerindeki yeni kodların %30'undan fazlasının yapay zekâ tarafından yazıldığını gururla beyan etmektedir. Ne var ki, 2024 tarihli Google Cloud DORA raporu, yapay zekâ benimsenmesindeki her %25'lik artışın sistemlerin teslimat istikrarında %7,2 ve yazılım teslimat hızında %1,5 oranında istatistiksel bir düşüşe yol açtığını çarpıcı bir şekilde göstermektedir. Üretim kapasitesi patlarken bunu denetleyecek insan kapasitesi yetersiz kalmakta ve “kendi sonunu kodlayan uzmanlar” paradoksu ortaya çıkmaktadır. Tecrübeli çalışanlar, kendi sezgilerini makinelere öğreterek paradoksal bir şekilde kendi yerlerini alacak sistemleri eğitmektedirler.

Bu krizin en ağır bedelini ise sisteme yeni girmeye çalışan gençler ödemektedir. Şirketler, giriş seviyesindeki ölçülebilir rutin işleri makinelere devrettikçe geleceğin uzmanlarını ve denetmenlerini yetiştirecek çıraklık altyapısı da çökmektedir. Görev ikamesi olarak bilinen bu olgu, sisteme yeni girenlerin bilgi birikimi edinme kanallarını tıkadığında yapısal bir krize dönüşmektedir. Şirketlerin kısa vadeli maliyet avantajı uğruna çıraklık altyapısını tahrip etmesi, güncel literatürde “kayıp genç döngüsü” (Catalini vd., 2026) olarak tanımlanmaktadır. Nitekim Brynjolfsson ve arkadaşlarının (2025) araştırması, yapay zekaya maruz kalan mesleklerde 22-25 yaş arası genç çalışanların istihdamının şimdiden %16 oranında düştüğünü göstermektedir. Sistemin ahlaki sorumluluktan yoksun bu verimlilik arayışı, kendi kendini yenileme kapasitesini yok eden bir yıkım motoruna dönüşmüştür.

Piketty'e (2020) göre hiçbir eşitsizlik rejimi sadece ekonomik zorbalıkla ayakta kalamaz ve kendini meşrulaştıracak ideolojik bir anlatıya ihtiyaç duyar. Silikon Vadisi'nin yapay zekâ ile gelen evrensel bolluk anlatısı, tam da bu türden bir meşrulaştırma aracıdır. Vitrindeki bu iyimser tablonun arkasında devasa bir uçurum yatmaktadır. Chopra ve arkadaşlarının (2025) geliştirdiği Buzdağı Endeksine göre ABD'de yapay zekânın ücretler üzerindeki ölçülebilir ve doğrudan etkisi %2,2 gibi zararsız bir seviyede görünürken gizli maruziyet ve tehdit altındaki ücret değeri aslında %11,7'ye ulaşmaktadır. Sermayeyi ve işlem gücünü elinde tutan teknoloji şirketleri, bu buzdağının altındaki kütleyi kendi lehlerine kullanırken eşitsizliği teknolojik bir kaçınılmazlık gibi sunarak ahlaki sorumluluktan ustaca sıyrılmaktadırlar. Öte yandan devletlerin de kıt kamu kaynaklarını sivil refahı artıracak doğrulama altyapılarına değil, giderek artan oranlarda güvenlik ve savunma sanayisine akıtması, bu eşitsizlik buzdağının köklerini daha da derinleştirmektedir.

Elbette teknoloji iktisadî literatüründe, savunma ve güvenlik odaklı Ar-Ge yatırımlarının tarihsel olarak internet veya GPS gibi devrimsel sivil teknolojilere zemin hazırladığı (yayılma etkisi) argümanı geçerliliğini korumaktadır. Benzer şekilde, serbest piyasa savunucuları, sistemin zamanla kendi iç denetim ve doğrulama mekanizmalarını inşa edeceğini ileri sürebilir. Ancak yapay zekânın karar alma otonomisi, emek piyasalarındaki yıkımın (örneğin kayıp genç döngüsünün) telafi edilemez hızı ve ölçülebilirlik açığının yapısal doğası, bu geleneksel piyasa iyimserliğini geçersiz kılmaktadır.

3. İnsanı yeniden denkleme katmak

Binlerce yıldır dünyadaki ilerlemenin ana motoru insan bilişi ve üretme kapasitesiydi. Ancak yapay zekâ ile bu tarihsel denklem tamamen tersine dönebilir. Yeni ekonomik kısıt zekâ veya üretim değil; insanlığın bu devasa ölçekteki üretimi denetleme, onaylama ve sorumluluğunu üstlenme kısıtıdır. Makinenin eşi benzeri görülmemiş üretim hızına karşılık insanın bilişsel ve bürokratik denetim hızının yavaş kalması, bu notun temel argümanlarından biri bağlamında “doğrulama bant genişliği daralması” olarak da tanımlanabilir. Bizler üretmekte değil, makinelerin hızla ürettiği ve optimize ettiği sonuçları doğrulamakta yavaş kalıyoruz. Hızla genişleyen bu ölçülebilirlik açığında, makineler daha otonom hale geldikçe Weber’in uyardığı o ruhsuz demir kafesin içine çekilme riskiyle karşı karşıyayız.

Weber’in demir kafesi ve Polanyi’nin toplumdan kopuk piyasa uyarıları, yapay zekâ çağında yalnızca felsefi birer tespit değil, acil kamu politikası müdahalesi gerektiren krizlerin habercisidir. Sistemin ahlaki sorumluluktan yoksun bu verimlilik arayışının yarattığı tahribat, piyasanın kendi dinamikleriyle çözölemeyecek kadar derindir.

Piyasayı toplumdan koparan ve bizi görünürde zengin ama temelde kırılgan bir içi boş ekonomiye sürükleyen bu gidişatı durdurmak için ihtiyaç duyduğumuz yeni sosyal sözleşme; soyut ahlaki çağrılarla değil, teorik tespitleri radikal ve somut düzenleyici mekanizmalara dönüştüren adımlarla inşa edilebilir.

Bugün teknoloji şirketleri, otomasyonun getirdiği kısa vadeli verimlilik kazanımlarını özelleştirirken doğrulanmamış sistemlerin yarattığı yıkıcı riskleri ve Truva Atı dışsallığını toplumsallaştırmaktadır. Bu piyasa başarısızlığını çözmenin yolu, ahlaki çağrılar yapmak değil; teknoloji şirketlerini yarattıkları riskleri içselleştirmeye zorlayacak katı sorumluluk rejimleri inşa etmektir. Bir makinenin ürettiği hatanın ve toplumsal maliyetin bedeli fiyatlandırılmadığı sürece, hiçbir şirket gönüllü olarak hızını kesip o ahlaki fren mekanizmasını kurmayacaktır.

Bu doğrultuda yasa yapıcılar; yüksek riskli otonom sistemleri piyasaya süren şirketler için zorunlu sorumluluk sigortaları ve risk ağırlıklı sermaye gereksinimleri getirmelidir. Emek piyasalarındaki yıkımı, özellikle de geleceğin uzmanlarını yetiştiren çıraklık sisteminin çöküşünü durdurmak için radikal bir kamu politikası adımı atılmalıdır. Catalini, Hui ve Wu’nun (2026) da vurguladığı gibi doğrulama altyapısı ve insanların risksiz bir şekilde tecrübe kazanmasını sağlayan sentetik pratik alanları, tıpkı eğitim, yollar veya sağlık sistemi gibi temel bir kamu malı olarak konumlandırılmalıdır. Devletler, sadece şirketlerin ürettiği kapalı sistemlere bel bağlamak yerine; bağımsız denetim enstitüleri kurmalı, algoritmik hataların kaydedildiği kamuya açık sonuç sicilleri oluşturmalı ve güvenliği kitle kaynaklı hale getirecek sivil Ar-Ge fonlarını devreye sokmalıdır. Kaybolan giriş seviyesi işlerin yerini alacak yeni nesil çıraklık sistemleri devlet tarafından desteklenmelidir.

Bununla birlikte devlet, kurumsal yanlışlıkları kırmak için sadece yasaklayıcı ve düzenleyici bir polis memuru rolüyle yetinmemelidir. Yeni bir sanayi politikası aktörü olarak sahneye çıkmalı; vergi teşvikleri ve kamu alımları gücünü kullanarak insan emeğini salt maliyet odaklı ikame eden vasat teknolojiler yerine, Acemoğlu ve Johnson’un da (2023) altını çizdiği “İnsan Odaklı Yapay Zekâ” paradigmasını kurumsallaştırmalıdır. Devlet, insanın üretim ve karar alma kapasitesini merkeze alan ve güçlendiren yapay zekâ yatırımlarını destekleyen öncü bir piyasa yapıcı olmalıdır. Ancak bu şekilde insan kapasitesinin çöktüğü bir düzenden kurtulup insanın bilişsel yeteneklerinin ve ahlaki yargılarının makinelerle birlikte büyüdüğü bir artırılmış ekonomiye geçiş yapabiliriz.

Sonuç olarak, Adam Smith'in duygudaşlık kavramı üzerine inşa ettiği ahlaki piyasa vizyonu, ruhsuz kod satırlarına ve optimizasyon metriklerine indirgenemez. İktisadi ve teknolojik gelişme, sadece işlem gücünün hızına terk edilemeyecek kadar hayati bir süreçtir. Yapay zekaya yön veren de nihayetinde insan olmalıdır; çünkü makinenin eşi benzeri görülmemiş bir hızla ürettiği bu yeni dünyayı, ancak insanın ahlaki değerleri, şüphesi ve sorumluluk bilinci yaşanabilir kılacaktır.

Kaynakça

- Acemoğlu, D., & Johnson, S. (2023). *Power and progress: Our thousand-year struggle over technology and prosperity*. PublicAffairs.
- Brynjolfsson, E., Chandar, B., & Chen, R. (2025). Canaries in the coal mine? Six facts about the recent employment effects of artificial intelligence. *Stanford Digital Economy Lab Working Paper*.
- Catalini, C., Hui, X., & Wu, J. (2026). Some Simple Economics of AGI. *arXiv preprint arXiv:2602.20946*.
- Chopra, A., Bhattacharya, S., Salvador, D., Paul, A., Wright, T., Garg, A., Ahmad, F., Schwarze, A. C., Raskar, R., & Balaprakash, P. (2025). The iceberg index: Measuring skills-centered exposure in the AI economy. *arXiv*.
- Google Cloud. (2024). *Accelerate state of DevOps report*. DORA.
- Greenblatt, R., Denison, C., Wright, B., Roger, F., MacDiarmid, M., Marks, S., ... & Hubinger, E. (2024). Alignment faking in large language models. *arXiv preprint arXiv:2412.14093*.
- Piketty, T. (2020). *Capital and ideology*. Harvard University Press.
- Polanyi, K. (1944). *The great transformation: The political and economic origins of our time*. Farrar & Rinehart.
- Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*. Penguin UK.
- Smith, A. (1759). *The theory of moral sentiments*. A. Millar.
- Weber, M. (1905). *The Protestant ethic and the spirit of capitalism*. Routledge.