

Artificial Intelligence Index Report 2024: Top 10 Takeaways

- AI Surpasses Humans in Some Tasks:** AI has outperformed humans in areas like image classification, visual reasoning, and English understanding. However, it still lags in more complex tasks such as competition-level mathematics, visual commonsense reasoning, and planning.
- Industry Dominates AI Research:** In 2023, the industry developed 51 notable machine learning models, compared to only 15 from academia. Additionally, 21 models emerged from industry-academia collaborations, marking a new high.
- Cutting-Edge Models Are More Expensive:** The cost of training state-of-the-art AI models has reached unprecedented levels. For instance, OpenAI's GPT-4 required an estimated \$78 million for training, while Google's Gemini Ultra cost \$191 million.
- The U.S. Leads in AI Model Development:** In 2023, U.S.-based institutions produced 61 notable AI models, far ahead of the European Union's 21 and China's 15.
- Lack of Standardization in Responsible AI Evaluations:** Research shows a significant gap in standardized responsible AI reporting. Leading companies like OpenAI, Google, and Anthropic use different benchmarks, making it challenging to systematically compare risks and limitations.
- Generative AI Investment Surges:** Despite a decline in overall private AI investment in 2023, funding for generative AI skyrocketed nearly eightfold to \$25.2 billion. Major players such as OpenAI, Anthropic, Hugging Face, and Inflection secured substantial funding.
- AI Boosts Worker Productivity:** Studies in 2023 highlighted that AI helps workers complete tasks faster and improves output quality. AI also shows potential in narrowing the skill gap between low- and high-skilled workers. However, improper use of AI can lead to decreased performance.
- AI Accelerates Scientific Progress:** In 2023, AI furthered scientific discovery with applications like AlphaDev for efficient algorithm sorting and GNoME for materials discovery.
- AI Regulations Increase in the U.S.:** The number of AI-related regulations in the U.S. grew significantly in 2023, reaching 25 compared to just one in 2016—a 56.3% rise in one year alone.
- Global Awareness and Concerns About AI Grow:** A survey by Ipsos found that 66% of people believe AI will dramatically impact their lives in the next 3–5 years, up from 60% in 2022. Additionally, 52% expressed nervousness about AI products, compared to 37% in 2022. In the U.S., Pew data shows a similar trend, with 52% feeling more concerned than excited about AI in 2023.

Frontier AI Research

- Frontier AI models include "notable models" (influential in AI research) and "foundation models" (large-scale AI models like GPT-4 and Gemini).
- Industry has dominated AI research since 2014. In 2023, 51 notable models were produced by the industry, while academia contributed 15.
- Collaboration between academia and industry resulted in 21 notable models in 2023, a record high.
- Training AI models requires immense computational resources and data, making them increasingly expensive.

Öncü Yapay Zeka Araştırmaları

- Öncü AI modelleri, "dikkate değer modeller" (AI araştırmalarında etkili olan) ve "temel modeller" (GPT-4 ve Gemini gibi büyük ölçekli AI modelleri) olarak tanımlanır.
- 2014'ten beri AI araştırmalarında endüstri hakim durumda. 2023'te 51 önemli model endüstri tarafından, 15 model ise akademi tarafından üretildi.
- Akademi ve endüstri arasındaki işbirlikleri, 2023'te 21 dikkate değer modelle rekor kırdı.

- AI modellerinin eğitimi büyük miktarda hesaplama kaynağı ve veri gerektiriyor ve bu da maliyetleri artırıyor.
-

Trends in AI Models

- **Geopolitical Distribution:**
 - The U.S. leads globally, producing 61 notable models in 2023, followed by China (15) and the European Union/UK combined (25).
 - Since 2019, the U.S. has consistently produced the highest number of AI models.
- **Model Complexity:**
 - Machine learning models have significantly increased in size and complexity, with parameter counts rising exponentially since the 2010s.
 - Models with higher parameters generally perform better but require more computational power and resources.
- **Training Compute:**
 - Computational demands for AI training have surged, with training compute usage increasing exponentially in the past decade.
 - Example: Google's Gemini Ultra required 50 billion petaFLOPs.

AI Modellerindeki Eğilimler

- **Jeopolitik Dağılım:**
 - ABD, 2023'te 61 dikkate değer modelle lider konumda, ardından Çin (15) ve AB/İngiltere (25) geliyor.
 - 2019'dan bu yana ABD, en fazla AI modeli üreten ülke konumunda.
 - **Model Karmaşıklığı:**
 - Makine öğrenimi modelleri 2010'lardan itibaren boyut ve karmaşıklık açısından önemli ölçüde büyümüştür.
 - Daha fazla parametreye sahip modeller genellikle daha iyi performans gösterir ancak daha fazla kaynak gerektirir.
 - **Eğitim İçin Hesaplama Gereksinimi:**
 - AI eğitimi için hesaplama gereksinimleri son 10 yılda hızla arttı.
 - Örneğin, Google'ın Gemini Ultra modeli 50 milyar petaFLOP gerektirdi.
-

Data Dependency

- Many foundational AI models rely on vast amounts of high-quality data from the internet.
- Concerns arise about "running out of data," especially for high-quality datasets by 2024 and image datasets by the 2040s.
- Synthetic data generation (AI-produced data) may address data shortages, but over-reliance risks "model collapse," where models lose the ability to generalize effectively.

Veriye Bağımlılık

- Temel AI modelleri, internetten elde edilen yüksek kaliteli verilere büyük ölçüde bağımlıdır.

- Özellikle 2024 itibarıyla yüksek kaliteli veri ve 2040'larda görsel veri sıkıntısı yaşanacağı endişesi bulunmaktadır.
 - Sentetik veri üretimi (AI tarafından oluşturulan veriler), veri kıtlığını çözebilir ancak aşırı bağımlılık, modellerin genelleme yapma yeteneğini kaybettiği "model çöküşü" riskini taşır.
-

Foundation Models

- Foundation models like GPT-4, Claude 3, and Llama 2 are versatile, trained on massive datasets, and used for various applications.
- Accessibility varies:
 - **Open access:** Fully available for modification (e.g., Meta's Llama 2).
 - **Limited access:** Access restricted via APIs (e.g., OpenAI's GPT-4).
 - **No access:** Exclusive to developers (e.g., Google's PaLM-E).
- In 2023, 65.8% of foundation models were open access, marking a shift toward openness in AI.

Temel Modeller

- GPT-4, Claude 3 ve Llama 2 gibi temel modeller çok yönlüdür, geniş veri setleri üzerinde eğitilir ve çeşitli uygulamalarda kullanılır.
 - Erişim çeşitlidir:
 - **Açık erişim:** Tamamen düzenlenebilir (ör. Meta'nın Llama 2 modeli).
 - **Sınırlı erişim:** API'ler aracılığıyla kısıtlı erişim (ör. OpenAI'nin GPT-4 modeli).
 - **Kapalı erişim:** Sadece geliştiricilere özel (ör. Google'ın PaLM-E modeli).
 - 2023'te temel modellerin %65.8'i açık erişimle sunuldu ve AI'de açıklığa doğru bir eğilim gösterdi.
-

Training Costs

- AI training costs have escalated drastically. For example:
 - The original Transformer in 2017 cost \$900 to train.
 - OpenAI's GPT-4 (2023) cost an estimated \$78 million.
 - Google's Gemini Ultra (2023) cost \$191 million.
- High training costs have excluded academia from creating state-of-the-art models, emphasizing the industry's dominance.

Eğitim Maliyetleri

- AI eğitim maliyetleri büyük ölçüde artmıştır. Örneğin:
 - 2017'deki orijinal Transformer modeli \$900'a eğitildi.
 - OpenAI'nin GPT-4 (2023) modeli tahmini \$78 milyon maliyetle eğitildi.
 - Google'ın Gemini Ultra (2023) modeli \$191 milyon maliyete ulaştı.
 - Yüksek eğitim maliyetleri, akademiye ileri düzey modeller geliştirme sürecinden dışladı ve endüstrinin hakimiyetini artırdı.
-

National and Organizational Leadership

- **National Affiliations:**
 - The U.S. produced 109 foundation models in 2023, followed by China (20) and the UK (8).
 - Since 2019, the U.S. has consistently led in originating the majority of foundation models.
- **Organizational Contributions:**
 - Industry leads with 72.5% of foundation models originating in 2023.
 - Google has released the most models (40 since 2019), followed by OpenAI (20).
 - Top academic contributors include Stanford and UC Berkeley.

Ulusal ve Kurumsal Liderlik

- **Ulusal Bağlantılar:**
 - ABD, 2023'te 109 temel model üretti; ardından Çin (20) ve İngiltere (8) geliyor.
 - 2019'dan bu yana ABD, çoğu temel modelin kaynağı olarak liderliğini sürdürüyor.
- **Kurumsal Katkıları:**
 - Endüstri, 2023'te üretilen temel modellerin %72.5'ini sağladı.
 - Google, 2019'dan bu yana 40 modelle en çok model üreten şirket oldu; ardından OpenAI (20 model) geliyor.
 - Akademik katkılar arasında Stanford ve UC Berkeley öne çıkıyor.

Artificial Intelligence Index Report 2024: Overview

English Summary

The document provides a comprehensive review of AI advancements in 2023, covering a wide range of topics:

1. **Significant Achievements:**
 - AI surpassed human performance in tasks like image classification, basic reading comprehension, and visual reasoning but still lags in competition-level mathematics and commonsense reasoning.
 - Multimodal AI models like GPT-4 and Gemini demonstrated flexibility in handling text, images, and audio.
2. **Benchmark Development:**
 - As AI models saturate existing benchmarks (e.g., ImageNet, SQuAD), researchers developed new challenges like SWE-bench (coding), HEIM (image generation), and MoCa (moral reasoning).
3. **AI Model Releases:**
 - 2023 saw key models such as GPT-4 Turbo, Llama 2, and Stable Diffusion v2, advancing language, image, and audio processing.
4. **Emerging Trends:**
 - Generative AI is increasingly used for creating synthetic datasets, accelerating training for future systems.
 - Human evaluation methods are gaining prominence over automated metrics.
5. **Robotics and Agents:**
 - Integration of LLMs with robotics, exemplified by models like PaLM-E, has improved robotic interactions and flexibility.

- Autonomous agents can now tackle complex tasks like Minecraft gameplay and online shopping.

6. Environmental Concerns:

- AI training and inference have substantial environmental impacts. Innovations like Direct Preference Optimization aim to minimize energy usage.

Türkçe Özet

Bu belge, 2023'teki yapay zeka gelişmelerini kapsamlı bir şekilde ele alıyor:

1. Önemli Başarılar:

- AI, görsel sınıflandırma, temel okuduğunu anlama ve görsel akıl yürütme gibi alanlarda insan performansını aştı. Ancak ileri düzey matematik ve sağduyu akıl yürütme gibi karmaşık görevlerde geride kaldı.
- GPT-4 ve Gemini gibi çok modlu modeller, metin, görsel ve ses işleme konularında esneklik sergiledi.

2. Benchmark Geliştirmeleri:

- Mevcut benchmark'ların (ImageNet, SQuAD) doygunluğa ulaşmasıyla araştırmacılar SWE-bench (kodlama), HEIM (görsel oluşturma) ve MoCa (ahlaki akıl yürütme) gibi yeni testler geliştirdi.

3. AI Model Çıktıları:

- GPT-4 Turbo, Llama 2 ve Stable Diffusion v2 gibi modeller, dil, görsel ve ses işleme alanlarında ilerleme sağladı.

4. Yeni Eğilimler:

- Üretken yapay zeka, gelecekteki sistemler için sentetik veri setleri oluşturarak eğitim süreçlerini hızlandırıyor.
- İnsan değerlendirme yöntemleri, otomatik ölçütlerin önüne geçiyor.

5. Robotik ve Ajanlar:

- Dil modellerinin robotik ile entegrasyonu, PaLM-E gibi modellerle robotların etkileşimlerini ve esnekliklerini artırdı.
- Otonom ajanlar artık Minecraft oynama ve çevrimiçi alışveriş gibi karmaşık görevleri yerine getirebiliyor.

6. Çevresel Endişeler:

- AI eğitimi ve çıkarımı, çevresel etkiler doğuruyor. Enerji kullanımını azaltmak için Direct Preference Optimization gibi yenilikler geliştiriliyor.

English Summary

The provided document focuses on fairness in artificial intelligence (AI), exploring challenges, advancements, and solutions in the context of responsible AI. Below is a detailed breakdown:

Challenges in AI Fairness

1. Defining and Measuring Fairness:

- There is no universal definition of fairness, making it challenging to implement across AI systems.
- Fairness needs to be integrated at every stage of the AI lifecycle.

2. Academia's Role:

- Increased awareness of fairness issues due to large language models (LLMs) like ChatGPT and Gemini.
- Submissions related to AI fairness and bias quadrupled between 2019 and 2023, with 212 submissions in 2023.

3. Industry's Perspective:

- A survey revealed that 29% of organizations consider fairness risks relevant to their AI strategies.
- European organizations lead in identifying fairness concerns (34%), compared to North America (20%).

Research Highlights

1. Fairness in Healthcare:

- Studies reveal racial bias in responses generated by LLMs like GPT-3.5 and GPT-4 to medical questions.
- Models perpetuate race-based misconceptions, such as differences in pain thresholds or lung capacities between races.

2. Social Bias in Image Generation:

- Tools like BiasPainter evaluate bias in models like DALL-E and Stable Diffusion.
- Bias was found most prevalent in age and race dimensions, with less bias in gender.

3. Subjective Opinions in LLMs:

- LLMs tend to align with Western perspectives, as shown in the GlobalOpinionQA dataset.
- This highlights the lack of diversity in model outputs, especially concerning non-Western opinions.

Solutions and Recommendations

1. Fairness in Model Design:

- Comprehensive fairness measures are essential, though the global average of adoption remains low (1.97 out of 5 measures).
- Organizations should prioritize integrating fairness in AI systems comprehensively.

2. Improved Tools and Metrics:

- Bias detection tools like BiasPainter offer better evaluation, achieving 90.8% accuracy in automatic bias detection tasks.
- The GlobalOpinionQA dataset serves as a benchmark for assessing opinion diversity in LLMs.

Türkçe Özet

Sağlanan belge, yapay zekada adalet kavramına odaklanmakta, sorumluluk sahibi yapay zeka bağlamında zorlukları, ilerlemeleri ve çözümleri ele almaktadır. İşte detaylı bir özet:

Yapay Zekada Adaletin Zorlukları

1. Adaletin Tanımlanması ve Ölçülmesi:

- Adaletin evrensel bir tanımı olmadığından, AI sistemlerinde uygulanması zor.

- Adalet, AI yaşam döngüsünün her aşamasına entegre edilmelidir.

2. Akademik Perspektif:

- ChatGPT ve Gemini gibi büyük dil modelleri sayesinde adalet konularına dair farkındalık arttı.
- 2019 ile 2023 arasında AI adaleti ve yanlılık ile ilgili akademik bildiri sayısı dört katına çıkarak 2023'te 212'ye ulaştı.

3. Endüstriyel Perspektif:

- Bir ankete göre, organizasyonların %29'u adalet risklerini AI stratejileri açısından önemli buluyor.
- Avrupa organizasyonları adalet konularında lider (%34), Kuzey Amerika ise daha düşük farkındalık gösteriyor (%20).

Araştırma Bulguları

1. Sağlık Alanında Adalet:

- GPT-3.5 ve GPT-4 gibi modellerin tıbbi sorulara verdiği yanıtlar ırksal yanlılıklar içeriyor.
- Modeller, ırk temelli yanlış algıları sürdürüyor (ör. ağrı eşiği farklılıkları).

2. Görsel Üretimde Sosyal Yanlılık:

- BiasPainter gibi araçlar, DALL-E ve Stable Diffusion gibi modellerdeki yanlılığı değerlendiriyor.
- Yaş ve ırk boyutlarında daha fazla yanlılık bulunurken, cinsiyet boyutunda daha az yanlılık görülmekte.

3. LLM'lerde Öznel Görüşler:

- LLM'ler, genellikle Batı perspektiflerine uygun yanıtlar üretiyor.
- GlobalOpinionQA veri seti, modellerin farklı bölgesel görüşleri yeterince temsil edemediğini gösteriyor.

Çözümler ve Öneriler

1. Model Tasarımında Adalet:

- Kapsamlı adalet önlemleri kritik önemde olsa da, küresel ortalama uygulama oranı düşüktür (5 üzerinden 1.97).
- Organizasyonlar, adaleti kapsamlı bir şekilde entegre etmeye öncelik vermelidir.

2. Geliştirilmiş Araçlar ve Ölçütler:

- BiasPainter gibi yanlılık tespit araçları, %90.8 doğrulukla daha iyi değerlendirme sunuyor.
- GlobalOpinionQA veri seti, LLM'lerde görüş çeşitliliğini değerlendirmek için bir ölçüt olarak kullanılıyor.

English Summary

The document focuses on the intersection of artificial intelligence (AI) and political processes, highlighting the risks, advancements, and mitigation strategies associated with AI in elections. Below are the key insights:

1. Deepfakes and Disinformation

- **Deepfake Capabilities:** AI generates realistic deepfakes, including face-swapping, voice manipulation, and entire face synthesis, which pose risks to political stability.

- **Disinformation Spread:** Tools like Countercloud automate disinformation pipelines, creating and disseminating fake articles and comments. Such systems cost as little as \$400 to set up.
- **Slovakia Case Study:** During the 2023 elections, an AI-manipulated audio clip influenced voter opinions, highlighting the role of deepfake disinformation in tight electoral contests.

2. Detecting Deepfakes

- Detection methods for AI-generated content struggle to generalize across new datasets.
- Accuracy in detecting deepfakes varies by demographic attributes, with the worst performance for darker skin tones.

3. Large Language Models (LLMs) and Bias

- LLMs like ChatGPT exhibit political bias, favoring parties such as the Democrats in the U.S. and Labour in the U.K.
- Bias in LLMs may subtly influence users' political views.

4. AI's Broader Impact on Politics

- AI applications include voter list maintenance, polling booth predictions, and voter authentication.
- Risks include biased algorithms, surveillance concerns, and potential to undermine electoral integrity.
- Mitigation strategies emphasize transparency, auditing, and bias reduction in AI tools.

5. Audio Deepfakes

- Audio deepfakes increasingly deceive people. Human detection accuracy is only 73%, and further advancements in generation methods may worsen detection rates.

Türkçe Özet

Belge, yapay zekanın (AI) siyasi süreçlerle kesişimini ele alarak, seçimlerde AI ile ilişkili riskleri, ilerlemeleri ve bunlara yönelik önleme stratejilerini vurgulamaktadır. İşte ana noktalar:

1. Deepfake ve Yanıltıcı Bilgiler

- **Deepfake Yetkinlikleri:** AI, yüz değiştirme, ses manipülasyonu ve tam yüz sentezi gibi gerçekçi deepfake içerikler üreterek siyasi istikrara tehdit oluşturuyor.
- **Yanlış Bilgi Yayılımı:** Countercloud gibi araçlar, yanlış bilgi oluşturma ve yayma süreçlerini otomatikleştiriyor. Bu sistemlerin kurulumu sadece 400 dolara mal oluyor.
- **Slovakya Örneği:** 2023 seçimlerinde, AI tarafından manipüle edilen bir ses kaydı seçmen görüşlerini etkiledi ve seçim sonuçlarını değiştirdi.

2. Deepfake Tespiti

- AI tarafından oluşturulan içeriklerin tespitinde kullanılan yöntemler yeni veri setlerinde genelleme yapmakta zorlanıyor.
- Tespit doğruluğu demografik özelliklere göre değişiyor; en düşük performans koyu ten rengine sahip kişilerde görülüyor.

3. Büyük Dil Modelleri (LLM) ve Yanlılık

- ChatGPT gibi LLM'ler, ABD'de Demokratlar ve İngiltere'de İşçi Partisi lehine politik yanlılık gösteriyor.
- Bu tür yanlılıklar, kullanıcıların siyasi görüşlerini ince bir şekilde etkileyebilir.

4. AI'nın Siyaset Üzerindeki Geniş Etkisi

- AI, seçmen listesi düzenleme, sandık yeri tahmini ve seçmen doğrulama gibi uygulamalarda kullanılıyor.
- Riskler arasında yanlış algoritmalar, gözetim endişeleri ve seçim bütünlüğünün zedelenmesi yer alıyor.
- Önleme stratejileri, şeffaflık, denetim ve yanlışlığın azaltılmasına odaklanıyor.

5. Ses Deepfakeleri

- Ses deepfakeleri, insanları giderek daha fazla yanıltıyor. İnsanların bu tür deepfakeleri tespit etme doğruluğu yalnızca %73 olup, daha iyi üretim yöntemleriyle bu oran daha da düşebilir.

English Summary

The document highlights major AI-related policy developments and governance initiatives in 2023, covering legislative efforts, global cooperation, and institutional advancements:

1. China's AI Regulations:

- Focus on "deep synthesis" technology to regulate realistic virtual entities and deepfake media.
- Measures include content moderation, user identity verification, and biometric editing consent.

2. United States Initiatives:

- **AI for National Security Act:** Enhances the Department of Defense's cyber-defense with AI tools.
- **AI Leadership Training Act:** Mandates AI literacy programs for federal leaders to promote ethical AI use.
- **Jobs of the Future Act:** Studies AI's impact on industries, skills, and employment.
- **National AI Commission Act:** Proposes a regulatory framework for AI to preserve U.S. leadership and mitigate risks.
- **Protect Elections from Deceptive AI Act:** Bans deceptive AI-generated political ads.

3. Voluntary Commitments by AI Firms:

- Seven major AI firms (e.g., Google, Microsoft, OpenAI) pledge to ensure AI safety, reliability, and transparency in collaboration with the U.S. government.

4. European Union's AI Act:

- Implements a risk-based framework classifying AI systems and enforcing transparency for generative AI like ChatGPT.
- Prohibits systems with unacceptable risks, such as behavioral manipulation.

5. United Kingdom's AI Safety Initiatives:

- Establishes the world's first **AI Safety Institute** for research and risk mitigation.
- Hosts the **AI Safety Summit**, resulting in the Bletchley Declaration, a global commitment to AI safety.

Türkçe Özet

Belge, 2023'teki önemli AI politikaları ve yönetim girişimlerini ele alıyor. Yasama çabaları, küresel işbirliği ve kurumsal gelişmeler öne çıkıyor:

1. Çin'in AI Düzenlemeleri:

- Gerçekçi sanal varlıklar ve deepfake medyayı düzenlemek için "derin sentez" teknolojisine odaklanıyor.
- İçerik denetimi, kullanıcı kimlik doğrulaması ve biyometrik düzenleme izni gibi önlemler içeriyor.

2. Amerika Birleşik Devletleri Girişimleri:

- **AI for National Security Act:** Savunma Bakanlığı'nın siber güvenliğini AI araçlarıyla güçlendiriyor.
- **AI Leadership Training Act:** Federal liderler için etik AI kullanımını teşvik eden eğitim programları oluşturuyor.
- **Jobs of the Future Act:** AI'nin endüstriler, yetenekler ve istihdam üzerindeki etkilerini inceliyor.
- **National AI Commission Act:** AI risklerini azaltmak ve ABD'nin liderliğini sürdürmek için düzenleyici bir çerçeve öneriyor.
- **Protect Elections from Deceptive AI Act:** AI ile oluşturulan yanıltıcı siyasi reklamları yasaklıyor.

3. **AI Şirketlerinin Gönüllü Taahhütleri:**

- Yedi büyük AI şirketi (ör. Google, Microsoft, OpenAI), güvenli, güvenilir ve şeffaf AI geliştirmeyi taahhüt ediyor.

4. **Avrupa Birliği AI Yasası:**

- AI sistemlerini sınıflandıran risk tabanlı bir çerçeve uygular.
- ChatGPT gibi üretken AI sistemleri için şeffaflık zorunluluğu getirir.
- Davranış manipülasyonunu yasaklayan sistemleri engeller.

5. **Birleşik Krallık AI Güvenlik Girişimleri:**

- Araştırma ve risk azaltma için dünyanın ilk **AI Güvenlik Enstitüsü** kuruldu.
- **AI Güvenlik Zirvesi**, küresel işbirliği sağlamak için Bletchley Deklarasyonu ile sonuçlandı.