

AI News Digest

03/16/2026 - 03/22/2026

WEEKLY DIGEST: OVERALL AI NEWS

AI Governance and Ethics Take Center Stage

Recent discussions in the AI landscape emphasize the critical need for equity, diversity, and inclusion (EDI) in AI development, as highlighted by new research initiatives and frameworks. The role of large language models (LLMs) is being re-evaluated, with a focus on their influence in shaping discourse and power dynamics. Additionally, the establishment of the Anthropic Institute aims to address the societal and economic challenges posed by advanced AI. As AI systems evolve, the importance of trust and ethical governance becomes increasingly apparent, particularly in high-stakes sectors like healthcare and public administration.

KEY FINDINGS

- ✓ A new Collection aims to integrate EDI principles into the AI lifecycle.
- ✓ LLMs are being framed as active participants in truth production rather than just flawed systems.
- ✓ The Anthropic Institute will focus on the risks and economic effects of advanced AI.

IMPLICATIONS

- Enhanced EDI frameworks could lead to more equitable AI outcomes.
- Reframing LLMs may shift regulatory approaches and public perception of AI.
- The establishment of dedicated research institutes signals a growing recognition of AI's societal impact.

1. Emerging roles and trends of equity, diversity, and inclusion in artificial intelligence

Category: Research

Departments: Computer Science, Political Science & Public Administration, Sociology & Anthropology, Educational Leadership

Topics: AI Ethics & Safety, Fairness in Recidivism Prediction, Value Pluralism in Generative AI, Justice-Oriented AI Framework, Equity, Diversity, and Inclusion in AI

Impact Score: 28.0

Rubric Scores

Research Relevance: 5

Educational Value: 4

Innovation / Novelty: 5

Practical Impact: 4

Interdisciplinary Potential: 5

Ethical / Policy Implications: 5

This article outlines a new Collection aimed at integrating equity, diversity, and inclusion (EDI) principles into the AI lifecycle through empirical and theoretical research. The Collection seeks to advance interdisciplinary studies that critically assess how AI systems impact various human identity axes, promote methodological innovations linking technical AI research with societal values, and present diverse global perspectives on EDI. It features 14 accepted papers that explore topics such as fairness in recidivism prediction, value pluralism in generative AI, and a multilevel framework for justice-oriented AI, emphasizing the need for context-sensitive approaches and broader discussions on power and representation in AI. The Collection reflects a shift from narrow bias mitigation to addressing systemic inequities and social transformation within AI systems.

<https://link.springer.com/article/10.1007/s00146-026-02969-z>

2. From 'objectivity' to obedience: LLMs as discourse, discipline, and power

Category: Policy & Ethics

Departments: Computer Science, Political Science & Public Administration, Sociology & Anthropology

Topics: AI Ethics, LLM Bias Analysis, Power Dynamics in AI, Discourse and Knowledge Production

Impact Score: 28.0

Rubric Scores

Research Relevance: 5

Educational Value: 4

Innovation / Novelty: 5

Practical Impact: 4

Interdisciplinary Potential: 5

Ethical / Policy Implications: 5

This paper critiques the prevailing discourse on AI bias by framing large language models (LLMs) as active participants in the production of truth and subject formation, rather than merely flawed predictive systems. Drawing on Michel Foucault's theories of power and discourse, the author argues that LLMs reproduce historical hierarchies through their probabilistic language generation, thereby influencing epistemic norms and structural power relations. The analysis highlights that AI bias is fundamentally linked to the ways knowledge and subjects are constituted, challenging the notion of AI as a neutral technology and emphasizing the need to understand its role in shaping societal narratives.

<https://link.springer.com/article/10.1007/s00146-026-02994-y>

3. From ethics washing to enforceable accountability: hospitals as hubs of digital constitutionalism

Category: Policy & Ethics

Departments: Nursing, Public Health Sciences, Political Science & Public Administration

Topics: AI Ethics & Safety, AI Governance Framework, Corporate Liability Mechanisms, Transparency Measures, Participatory Governance

Impact Score: 27.0

Rubric Scores

Research Relevance: 4

Educational Value: 4

Innovation / Novelty: 5

Practical Impact: 4

Interdisciplinary Potential: 5

Ethical / Policy Implications: 5

The article proposes a comprehensive framework to enhance the enforceability of AI governance in healthcare, emphasizing the need for institutional empowerment, corporate liability, transparency, and participatory governance. It advocates for AI Ethics Boards to have conditional veto powers to halt non-compliant deployments and suggests embedding corporate responsibility mechanisms based on strict liability to ensure accountability for AI failures. The framework also highlights the importance of transparency measures, such as public registries and audit trails, to foster compliance and trust, while recommending that hospitals serve as the primary organizational locus for implementing these governance principles. Ultimately, the article calls for a shift in oversight from advisory roles to an institutional architecture capable of generating binding effects in AI governance.

<https://link.springer.com/article/10.1007/s00146-026-02980-4>

4. Pluralistic AI governance: Indigenous knowledge and the right of Indigenous peoples to self-determination

Category: Policy & Ethics

Departments: Political Science & Public Administration, Sociology & Anthropology, Educational Leadership

Topics: AI Policy & Regulation, Indigenous Knowledge Systems, Free, Prior and Informed Consent, AI Governance Frameworks, Capacity-Building for Indigenous Communities

Impact Score: 27.0

Rubric Scores

Research Relevance: 4

Educational Value: 4

Innovation / Novelty: 5

Practical Impact: 4

Interdisciplinary Potential: 5

Ethical / Policy Implications: 5

The article critiques global AI governance frameworks, particularly the proposed International AI Organization (IAIO), for marginalizing Indigenous participation and perpetuating hierarchical power dynamics. It argues that Indigenous peoples should not be treated merely as stakeholders but recognized as distinct entities with unique governance systems and knowledge. To achieve genuine inclusion, the IAIO must ensure ongoing Indigenous representation and authority, grounded in principles such as Free, Prior and Informed Consent (FPIC) and the CARE Principles. Additionally, the article emphasizes the need for dedicated funding and capacity-building to empower Indigenous communities in AI governance, moving beyond tokenism to substantive engagement.

<https://link.springer.com/article/10.1007/s00146-026-02938-6>

5. New Anthropic Institute to Study Risks and Economic Effects of Advanced AI

Category: Research
Departments: Political Science & Public Administration, Economics & Finance, Computer Science, Sociology & Anthropology
Topics: AI Policy & Regulation, Societal Impact Analysis, Economic Effects of AI, Red-Teaming in AI

Impact Score: 27.0

Rubric Scores

Research Relevance: 5	Educational Value: 4	Innovation / Novelty: 4
Practical Impact: 5	Interdisciplinary Potential: 4	Ethical / Policy Implications: 5

Anthropic has established the Anthropic Institute, a new research initiative aimed at investigating the societal, economic, and legal challenges posed by advanced AI systems. Led by co-founder Jack Clark, the institute will consolidate existing research efforts in red-teaming, societal impacts, and economic analysis, while also expanding into new areas. Key research questions will include the effects of powerful AI on jobs, economic activity, and governance, as well as the potential risks associated with these technologies. The institute plans to engage with affected communities and publish findings to inform both public discourse and the development of AI systems.

<https://campustechnology.com/Articles/2026/03/16/New-Anthropic-Institute-to-Study-Risks-and-Economic-Effects-of-Advanced-AI.aspx>

6. OpenAI is throwing everything into building a fully automated researcher

Category: Research
Departments: Computer Science, Biological Sciences, Chemistry & Biochemistry, Physics, Mathematical Sciences
Topics: Large Language Models, Reasoning Models, Automated Research, Scientific Problem Solving

Impact Score: 26.0

Rubric Scores

Research Relevance: 5	Educational Value: 4	Innovation / Novelty: 5
Practical Impact: 5	Interdisciplinary Potential: 4	Ethical / Policy Implications: 3

Researchers at OpenAI are advancing the capabilities of large language models (LLMs) to tackle long-running scientific research tasks, building on the success of coding agents like Codex. The development of reasoning models has enhanced the ability of these systems to work independently for extended periods, enabling them to manage complex problems by breaking them into subtasks. Recent applications of GPT-5 have led to new solutions for unsolved mathematical problems and breakthroughs in biology, chemistry, and physics. OpenAI aims to apply the problem-solving capabilities demonstrated in coding to broader scientific challenges, prioritizing real-world relevance over theoretical achievements.

<https://www.technologyreview.com/2026/03/20/1134438/openai-is-throwing-everything-into-building-a-fully-automated-researcher/>

7. AI-powered robot learns how to harvest tomatoes more efficiently

Category: Research

Departments: Industrial, Manufacturing & Systems Engineering, Computer Science, Electrical & Computer Engineering

Topics: Robotics, Harvest-Ease Estimation, Image Recognition, Agricultural Robotics

Impact Score: 26.0

Rubric Scores

Research Relevance: 5

Educational Value: 4

Innovation / Novelty: 5

Practical Impact: 5

Interdisciplinary Potential: 4

Ethical / Policy Implications: 3

Assistant Professor Takuya Fujinaga from Osaka Metropolitan University has developed a robotic system that enhances tomato harvesting by assessing the ease of picking each fruit. This system integrates image recognition and statistical analysis to evaluate factors such as the tomato's visibility and position relative to stems and leaves, leading to a new metric termed "harvest-ease estimation." In testing, the system achieved an 81% success rate, demonstrating its ability to adapt its approach when initial attempts fail. The research aims to advance agricultural robotics by enabling robots to make informed decisions, potentially transforming human-robot collaboration in farming.

<https://www.sciencedaily.com/releases/2026/03/260317064512.htm>

8. Measuring progress toward AGI: A cognitive framework

Category: Research

Departments: Computer Science, Psychology, Engineering Education & Leadership

Topics: Artificial General Intelligence, Cognitive Taxonomy, Evaluation Methods, Cognitive Abilities

Impact Score: 26.0

Rubric Scores

Research Relevance: 5

Educational Value: 4

Innovation / Novelty: 5

Practical Impact: 4

Interdisciplinary Potential: 5

Ethical / Policy Implications: 3

The article announces the release of a paper titled "Measuring Progress Toward AGI: A Cognitive Taxonomy," which aims to establish a scientific framework for evaluating the cognitive capabilities of AI systems as they progress toward Artificial General Intelligence (AGI). The framework is based on research from psychology, neuroscience, and cognitive science, identifying ten key cognitive abilities deemed essential for general intelligence in AI. Additionally, the authors are collaborating with Kaggle to host a hackathon, encouraging the research community to develop evaluation methods that implement this cognitive taxonomy.

<https://deepmind.google/blog/measuring-progress-toward-agi-a-cognitive-framework/>

9. Reciprocal trust and distrust in artificial intelligence systems: the hard problem of regulation

Category: Policy & Ethics

Departments: Political Science & Public Administration, Computer Science, Public Health Sciences

Topics: AI Ethics & Safety, Trust in AI Systems, Watchful Trust, AI Governance, High-Stakes AI Applications

Impact Score: 26.0

Rubric Scores

Research Relevance: 4

Educational Value: 4

Innovation / Novelty: 4

Practical Impact: 5

Interdisciplinary Potential: 4

Ethical / Policy Implications: 5

The article explores the dynamics of trust and distrust in the context of human-AI interactions, particularly in high-stakes domains such as public administration and healthcare. It emphasizes that trust is a relational belief based on expectations of competence and goodwill, while distrust can serve as a beneficial force when it promotes oversight and accountability. The concept of "watchful trust" is introduced, advocating for a balanced approach where trust is contingent on demonstrated trustworthiness, particularly in AI systems, which must meet ethical and technical standards to be deemed trustworthy. Empirical evidence indicates that the alignment of trust levels significantly impacts decision quality in public-sector applications, highlighting the need for careful management of trust in AI governance.

<https://link.springer.com/article/10.1007/s00146-026-02966-2>

10. Scientists discover AI can make humans more creative

Category: Research

Departments: Educational Leadership, Computer Science, Psychology

Topics: Generative AI, AI-Enhanced Creativity, Diversity in AI Outputs, User Engagement in Design

Impact Score: 25.0

Rubric Scores

Research Relevance: 5

Educational Value: 4

Innovation / Novelty: 5

Practical Impact: 4

Interdisciplinary Potential: 4

Ethical / Policy Implications: 3

Research from Swansea University challenges the conventional view of AI as merely a tool for automating tasks, presenting it instead as a creative collaborator that enhances human engagement and exploration. In a study involving over 800 participants, an AI-supported system utilized the MAP-Elites method to generate diverse design galleries for virtual cars, leading to improved design outcomes and increased user involvement. The findings suggest that traditional metrics for evaluating AI tools may be inadequate, as they often fail to capture the broader impacts on creativity and user experience. The study emphasizes the importance of diversity in AI-generated outputs, indicating that a range of ideas, including flawed ones, can stimulate creative thinking and prevent fixation on initial concepts.

<https://www.sciencedaily.com/releases/2026/03/260315004355.htm>