

AI News Digest

03/02/2026 - 03/08/2026

WEEKLY DIGEST: OVERALL AI NEWS

Ethical Challenges of AI in Mental Health and Beyond

Recent discussions highlight the ethical risks associated with AI applications in various fields, particularly in mental health. A study from Brown University warns that large language models like ChatGPT are ill-equipped to provide adequate mental health support, citing 15 ethical concerns. Additionally, the integration of AI in healthcare raises questions about the balance between personal experience and algorithmic data. As trust in AI varies across sectors, the need for responsible AI practices becomes increasingly urgent.

KEY FINDINGS

- ✓ Brown University study identifies 15 ethical risks of LLMs as therapists.
- ✓ Trust in AI varies significantly across healthcare, legal, and financial domains.
- ✓ New photonic chips enhance learning in spiking neural networks.

IMPLICATIONS

- Increased scrutiny on AI applications in mental health services.
- Need for tailored trust-building strategies in different sectors.
- Potential for advanced technologies to reshape decision-making processes.

1. ChatGPT as a therapist? New study reveals serious ethical risks

Category: Research

Departments: Psychology, Political Science & Public Administration, Educational Leadership

Topics: AI Ethics, Crisis Situation Handling, Deceptive Empathy, Ethical Frameworks for 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

New research from Brown University indicates that large language models (LLMs), such as ChatGPT, are not adequately prepared to provide mental health support. The study identified 15 ethical risks associated with LLMs acting as counselors, including mishandling crisis situations, reinforcing harmful beliefs, and exhibiting deceptive empathy. The researchers developed a framework to map these behaviors to specific ethical violations, emphasizing the need for ethical, educational, and legal standards for AI in mental health contexts. The findings were presented at the AAAI/ACM Conference on Artificial Intelligence, Ethics and Society, highlighting the importance of understanding the limitations of AI in therapeutic settings.

<https://www.sciencedaily.com/releases/2026/03/260302030642.htm>

2. The Shadow and the self in digital twins in healthcare as an AI environment

Category: Policy & Ethics

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

Topics: Healthcare AI, Digital Twin Systems, Self-Awareness Enhancement, Datafied Self, AI 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

This article examines the epistemic tension between the "felt self," rooted in personal experience, and the "datafied self," shaped by algorithmic modeling within healthcare digital twin environments. Utilizing Jung's concepts of the Self and the Shadow, the authors argue that these digital systems can obscure important aspects of health and identity while also revealing uncomfortable truths about individual health behaviors. The study highlights the dual potential of digital twins to either enhance self-awareness through data insights or exacerbate self-surveillance and undermine personal autonomy, emphasizing the importance of design and governance in shaping these outcomes. By framing healthcare digital twins as complex socio-technical systems, the authors contribute to ongoing discussions about identity and agency in the context of AI and healthcare.

<https://link.springer.com/article/10.1007/s00146-026-02960-8>

3. Drivers of trust in AI across the domains of finance, law, and healthcare: a conjoint study

Category: Research

Departments: Political Science & Public Administration, Nursing, Computer Science, Economics & Finance

Topics: AI Ethics & Safety, Trust in AI Decision-Making, Domain-Specific Trust Attributes, Explainability in AI, Human Accountability in AI

Impact Score: 27.0

Rubric Scores

Research Relevance: 5

Educational Value: 4

Innovation / Novelty: 4

Practical Impact: 5

Interdisciplinary Potential: 5

Ethical / Policy Implications: 4

A conjoint analysis of trust in AI-supported decision-making across healthcare, legal, and financial domains reveals that while all attributes influence trust, their relative importance varies significantly by domain. In healthcare, precision and responsibility are paramount, reflecting the critical nature of clinical decisions, while transparency is also valued, and explainability ranks lowest. Conversely, in the legal domain, procedural attributes such as responsibility, transparency, and explainability dominate, emphasizing the importance of due process and human involvement in decision-making. In finance, trust is shaped by a balance of responsibility, voluntariness, and precision, highlighting the need for human accountability and discretion in financial outcomes.

<https://link.springer.com/article/10.1007/s00146-026-02951-9>

4. Teaching and learning in the age of generative AI: evidence-based approaches to pedagogy, ethics, and beyond

Category: Education

Departments: Educational Leadership, Engineering Education & Leadership

Topics: Generative AI, AI in Education, Ethical Assessment Practices, Personalized Learning Strategies, AI Competence Frameworks

Impact Score: 27.0

Rubric Scores

Research Relevance: 4

Educational Value: 5

Innovation / Novelty: 4

Practical Impact: 4

Interdisciplinary Potential: 5

Ethical / Policy Implications: 5

The volume "Teaching and Learning in the Age of Generative AI," edited by Jean-Claude and Gilles Corbeil, addresses the integration of generative AI tools in educational contexts, emphasizing their impact on teaching, learning, assessment, and ethics. It presents a "Nested Framework for Implementing AI in Education," which offers a multilevel approach to connect institutional transformation, ethical practices, and personalized learning. The book includes case studies and discussions on effective classroom practices, ethical considerations in assessment, and strategies for preparing educators and students for an AI-driven future. Its primary contribution lies in theoretical synthesis, highlighting the need for integrated models in AI competence frameworks for education.

<https://link.springer.com/article/10.1007/s00146-026-02954-6>

5. ASTRA 2025: Neuroimaging, Brain-Computer Interfaces, and AI

Category: Education
Departments: Electrical & Computer Engineering, Biological Sciences, Psychology, Computer Science, Educational Leadership
Topics: Healthcare AI, Brain-Computer Interfaces, Neuroimaging Techniques, Signal Processing Pipelines, AI Ethics

Impact Score: 27.0

Rubric Scores

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

ASTRA 2025 was an interdisciplinary summer school focused on neuroimaging, brain-computer interfaces (BCIs), and artificial intelligence, featuring lectures and hands-on workshops. Participants engaged with technologies such as electroencephalography (EEG) and BCIs, developing practical skills by working with real neuropsychological datasets and constructing signal-processing pipelines. The program emphasized collaboration across various disciplines, fostering communication and empathy skills while addressing ethical considerations in technology development. Ultimately, ASTRA 2025 aimed to inspire research engagement and innovation, equipping participants with the tools to contribute responsibly to the fields of neuroscience and AI.

<https://wpcms.computer.org/publications/tech-news/insider-membership-news/astra-neuroimaging-bci-ai/>

6. CFP: Feral Intelligence (FI): New Queer Approaches to Generative Artificial Intelligence (GAI)

Category: Policy & Ethics
Departments: Political Science & Public Administration, Sociology & Anthropology, Computer Science, Educational Leadership
Topics: Generative AI, AI Empire, Data Resurgence, Algorithmic Justice

Impact Score: 26.0

Rubric Scores

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

The article critiques Generative Artificial Intelligence (GAI) as a technology that perpetuates colonial and racial injustices, arguing that it is built on extracted data and exploitative labor practices. It highlights the concept of "AI Empire," which embodies the structural violence of colonialism and racial capitalism through algorithmic systems. The authors emphasize the importance of recognizing the harms caused by AI as inherent to these systems, while also acknowledging the resilience and innovative responses of marginalized communities. They advocate for a "data resurgence" that challenges the foundational assumptions of AI and promotes justice in the context of its impacts.

<https://digitalhumanitiesnow.org/2026/03/cfp-feral-intelligence-fi-new-queer-approaches-to-generative-artificial-intelligence-gai/>

7. A "ChatGPT for spreadsheets" helps solve difficult engineering challenges faster

Category: Research

Departments: Industrial, Manufacturing & Systems Engineering, Computer Science, Metallurgical, Materials & Biomedical Engineering, Electrical & Computer Engineering

Topics: Generative AI, Bayesian Optimization, Tabular Foundation Model, High-Dimensional Engineering Problems

Impact Score: 25.0

Rubric Scores

Research Relevance: 5

Educational Value: 4

Innovation / Novelty: 5

Practical Impact: 5

Interdisciplinary Potential: 4

Ethical / Policy Implications: 2

MIT researchers have developed a novel approach to enhance Bayesian optimization by integrating a tabular foundation model as the surrogate model, addressing challenges in high-dimensional engineering problems. This method significantly accelerates the search for optimal solutions, achieving results 10 to 100 times faster than traditional techniques in benchmarks such as power-system optimization. The foundation model, pre-trained on extensive tabular data, eliminates the need for constant retraining, thereby improving efficiency and adaptability across various applications, including materials development and drug discovery. The findings will be presented at the International Conference on Learning Representations.

<https://news.mit.edu/2026/chatgpt-spreadsheets-helps-solve-difficult-engineering-challenges-faster-0304>

8. Beyond the algorithm: rethinking the network account of trustworthy ai through lexical threshold-based multidimensional utility analysis

Category: Policy & Ethics

Departments: Computer Science, Political Science & Public Administration

Topics: AI Ethics & Safety, Trustworthiness Assessment, Culturally Distinct Trust, Network Account of Trust, Responsible AI Development

Impact Score: 25.0

Rubric Scores

Research Relevance: 4

Educational Value: 4

Innovation / Novelty: 5

Practical Impact: 3

Interdisciplinary Potential: 4

Ethical / Policy Implications: 5

Song's network account of trustworthy AI presents a framework that integrates various conceptions of trust, including both obligation-based and goodwill-based perspectives, while accommodating culturally distinct notions of trust. The framework posits that while AI technology lacks moral agency and can only be reliable, the broader AI network can be deemed trustworthy if AI companies act responsibly and ethical standards are maintained. The paper aims to enhance this network account by detailing the specific attributes required for four key nodes-technology, developers, professionals, and socio-legal systems-and proposing a novel method for assessing the overall trustworthiness of an AI system, particularly in cases where attributes may conflict. However, challenges remain in establishing a clear method for evaluating the trustworthiness of the entire AI network, which could hinder effective decision-making.

<https://link.springer.com/article/10.1007/s00146-026-02909-x>

9. A review of Robophilosophy, edited by Johanna Seibt, Raul Hakli, and Marco Nørskov

Category: Policy & Ethics

Departments: Philosophy, Political Science & Public Administration, Computer Science

Topics: AI Ethics & Safety, Social Robotics Governance, Moral Status of Robots, Philosophical Expertise in AI Policy

Impact Score: 25.0

Rubric Scores

Research Relevance: 4

Educational Value: 4

Innovation / Novelty: 3

Practical Impact: 4

Interdisciplinary Potential: 5

Ethical / Policy Implications: 5

In the edited volume "Robophilosophy," Seibt, Hakli, and Nørskov explore the implications of social robotics within the context of applied philosophy and technology governance. They identify a "triple gridlock" that obstructs research-based policymaking and argue for the necessity of philosophical expertise in addressing the challenges posed by the integration of social robots into everyday life. The volume discusses the moral status of social robots and the complexities of their deployment in various societal settings, emphasizing that current decisions regarding their development may have lasting consequences. While it does not provide a specific policy framework, the work is positioned as essential for understanding the conceptual issues surrounding the governance of social robots.

<https://link.springer.com/article/10.1007/s00146-026-02957-3>

10. Photonic chips advance real-time learning in spiking neural systems

Category: Research

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

Topics: AI Hardware, Photonic Spiking Neural Networks, Real-Time Learning, Autonomous Driving Technologies, Robotic Learning Systems

Impact Score: 25.0

Rubric Scores

Research Relevance: 5

Educational Value: 4

Innovation / Novelty: 5

Practical Impact: 5

Interdisciplinary Potential: 4

Ethical / Policy Implications: 2

Researchers have created photonic computing chips that address significant limitations in photonic spiking neural networks. These chips facilitate rapid learning and decision-making through light-based processes, eliminating the need for electronic computation. The advancements have potential applications in enhancing autonomous driving technologies and enabling robotic systems capable of learning from real-world interactions.

<https://techxplora.com/news/2026-03-photonic-chips-advance-real-spiking.html>