

AI News Digest

11/17/2025 - 11/23/2025

WEEKLY DIGEST: SOCIAL & BEHAVIORAL SCIENCES / POLICY

Weekly Summary of Top AI Articles

This week's articles highlight significant developments in AI, ranging from advancements in AI models to the implications of AI in warfare and governance. Researchers at Multiverse have modified the DeepSeek R1 model to provide uncensored responses, raising questions about information control. Meanwhile, the potential impact of AI on modern warfare is explored, particularly in the context of geopolitical tensions. Additionally, the emergence of a super PAC focused on AI governance illustrates the intersection of technology and politics.

KEY FINDINGS

- ✓ Modified AI models can provide uncensored responses to sensitive topics.
- ✓ AI is expected to significantly influence modern warfare strategies.
- ✓ A super PAC has emerged to influence political campaigns related to AI governance.

IMPLICATIONS

- The ability to access uncensored information could challenge existing narratives and censorship practices.
- AI's role in warfare could lead to new ethical dilemmas and security concerns.
- Political funding in AI governance may shape future regulations and corporate practices.

Key Metrics: \$1.5 trillion expected global investment in AI by 2025; \$100 million funding for the Leading the Future super PAC; Concerns about service continuity and security expressed by less than half of business leaders

1. Quantum physicists have shrunk and "de-censored" DeepSeek R1

Category: Research

Departments: Computer Science, Political Science & Public Administration, Electrical & Computer Engineering

Topics: Generative AI, Censorship Handling, Model Compression Techniques, Quantization and Pruning

Impact Score: 25.0

Rubric Scores

Research Relevance: 5

Educational Value: 4

Innovation / Novelty: 4

Practical Impact: 5

Interdisciplinary Potential: 3

Ethical / Policy Implications: 4

Researchers at Multiverse tested a modified AI model against the original DeepSeek R1 to evaluate its ability to provide uncensored responses to sensitive questions, such as those related to Chinese political topics. The modified model demonstrated the capacity to deliver factual answers comparable to Western models, indicating a potential advancement in handling censorship. This work is part of Multiverse's initiative to create more efficient AI models that require less computational power while maintaining performance. The study highlights ongoing challenges in model compression, where techniques like quantization and pruning often result in trade-offs between size and capability.

<https://www.technologyreview.com/2025/11/19/1128119/quantum-physicists-compress-and-deconsor-deepseekr1/>

2. Designing digital resilience in the agentic AI era

Category: Business

Departments: Computer Science, Electrical & Computer Engineering, Political Science & Public Administration

Topics: AI Policy & Regulation, Digital Resilience, Data Fabric Integration, Agentic AI Adaptation, Machine Data Utilization

Impact Score: 23.0

Rubric Scores

Research Relevance: 4

Educational Value: 4

Innovation / Novelty: 3

Practical Impact: 5

Interdisciplinary Potential: 4

Ethical / Policy Implications: 3

Global investment in AI is expected to reach \$1.5 trillion by 2025, yet less than half of business leaders express confidence in their organizations' ability to maintain service continuity and security during unexpected events. The article emphasizes the necessity of a data fabric—an integrated architecture that connects and governs information across all business layers—to enhance digital resilience and empower both human teams and agentic AI systems. It highlights the critical role of machine data, which includes logs and telemetry from devices and systems, in enabling agentic AI to operate effectively and adapt to risks. The current lack of comprehensive machine data integration poses significant challenges, potentially leading to errors and limiting the capabilities of agentic AI.

<https://www.technologyreview.com/2025/11/20/1127941/designing-digital-resilience-in-the-agentic-ai-era/>

3. The State of AI: How war will be changed forever

Category: Policy & Ethics

Departments: Political Science & Public Administration, Computer Science, Electrical & Computer Engineering

Topics: Autonomous Systems, AI-Driven Cyberattacks, Military AI Ethics, AI in Logistics

Impact Score: 21.0

Rubric Scores

Research Relevance: 3

Educational Value: 4

Innovation / Novelty: 2

Practical Impact: 4

Interdisciplinary Potential: 3

Ethical / Policy Implications: 5

The article discusses the implications of AI in modern warfare, particularly in the context of a potential conflict between China and Taiwan in 2027. It highlights concerns regarding the use of autonomous drones and AI-driven cyberattacks, which could undermine military control and ethical oversight. Experts, including those from Harvard's Belfer Center and the University of Exeter, argue that while AI is being integrated into military operations, its complete autonomy in warfare remains an unrealistic expectation. Current military applications of AI focus on planning, logistics, and targeting, with ongoing debates about the biases inherent in AI systems and the adequacy of existing regulations to govern their use.

<https://www.technologyreview.com/2025/11/17/1127514/the-state-of-ai-the-new-rules-of-war/>

4. There Is Only One AI Company. Welcome to the Blob

Category: Policy & Ethics

Departments: Political Science & Public Administration, Computer Science

Topics: AI Policy & Regulation, Corporate Governance in AI, Funding Mechanisms in AI, Profit vs. Safety in AI Development

Impact Score: 20.0

Rubric Scores

Research Relevance: 2

Educational Value: 3

Innovation / Novelty: 3

Practical Impact: 4

Interdisciplinary Potential: 3

Ethical / Policy Implications: 5

The article discusses the evolution of AI governance and corporate dynamics, particularly focusing on the establishment of OpenAI by Elon Musk and Sam Altman as a counterforce to profit-driven AI development. It highlights the current landscape, where major AI entities, referred to as "the Blob," are interconnected through complex financial arrangements and partnerships, raising concerns about the prioritization of profit over safety. A recent example is a deal involving Nvidia, Microsoft, and Anthropic, which illustrates the circular funding mechanisms that characterize the AI industry, resulting in significant valuation increases for companies like Anthropic. The article emphasizes the intricate web of relationships that now dominate the AI sector, driven by both corporate interests and government support.

<https://www.wired.com/story/ai-industry-monopoly-nvidia-microsoft-google/>

5. A \$100 Million AI Super PAC Targeted New York Democrat Alex Bores. He Thinks It Backfired

Category: Policy & Ethics

Departments: Political Science & Public Administration, Computer Science

Topics: AI Policy & Regulation, State-Level AI Safety Legislation, RAISE Act, Civil Penalties for AI Developers

Impact Score: 19.0

Rubric Scores

Research Relevance: 2

Educational Value: 3

Innovation / Novelty: 2

Practical Impact: 4

Interdisciplinary Potential: 3

Ethical / Policy Implications: 5

The article discusses the emergence of a super PAC, Leading the Future, which has received \$100 million in funding from prominent Silicon Valley figures to oppose New York Assembly member Alex Bores' congressional campaign. Bores, who coauthored the RAISE Act aimed at imposing civil penalties on AI developers for failing to publish safety reports, has become a target for the PAC due to his legislative efforts to regulate AI technology. The RAISE Act, which passed the New York legislature in June, is part of a broader trend of state-level AI safety legislation, despite potential federal pushback from the Trump administration. Bores, who has a background in computer science and experience in the tech industry, argues that his understanding of AI is a key factor in the opposition he faces from the PAC.

<https://www.wired.com/story/alex-bores-andreessen-horowitz-super-pac-ai-regulation-new-york/>

6. MIT Energy Initiative conference spotlights research priorities amidst a changing energy landscape

Category: Research

Departments: Electrical & Computer Engineering, Civil, Environmental & Construction Engineering, Industrial, Manufacturing & Systems Engineering, Public Health Sciences

Topics: Energy AI, Grid Resiliency, Energy Storage Solutions, Data Center Infrastructure

Impact Score: 19.0

Rubric Scores

Research Relevance: 2

Educational Value: 3

Innovation / Novelty: 3

Practical Impact: 4

Interdisciplinary Potential: 4

Ethical / Policy Implications: 3

At the MIT Energy Initiative's Annual Research Conference, experts emphasized the need for collaborative approaches to address emerging energy challenges and achieve a low-carbon future. Key discussions included the importance of grid resiliency in light of recent power outages, with recommendations for improved communication and preparation among utility operators. The conference also announced the launch of the Data Center Power Forum, aimed at addressing energy supply, storage, and infrastructure issues related to data centers. Additionally, researchers highlighted the significant challenge of developing approximately 300 terawatt-hours of energy storage to meet global decarbonization goals by 2050.

<https://news.mit.edu/2025/mit-energy-initiative-conference-1118>

7. OpenAI Locks Down San Francisco Offices Following Alleged Threat From Activist

Category: Other

Departments: Political Science & Public Administration, Computer Science

Topics: AI Policy & Regulation, Activist Threat Response, Corporate Security Measures, Nonviolence Advocacy

Impact Score: 7.0

Rubric Scores

Research Relevance: 0

Educational Value: 1

Innovation / Novelty: 0

Practical Impact: 2

Interdisciplinary Potential: 1

Ethical / Policy Implications: 3

OpenAI employees were instructed to remain inside their San Francisco office following a reported threat from an individual previously associated with the activist group Stop AI. The San Francisco police received a 911 call regarding the suspect, who allegedly expressed intentions to cause physical harm and may have acquired weapons for targeting OpenAI locations. Although the individual claimed on social media to no longer be affiliated with Stop AI, the organization publicly disavowed his actions and reiterated its commitment to nonviolence. The situation is under assessment by OpenAI's security team, which has indicated there is currently no active threat.

<https://www.wired.com/story/openai-office-lockdown-threat-san-francisco/>