AI News Digest

10/05/2026 - 10/11/2026

WEEKLY DIGEST: HUMANITIES & CULTURAL STUDIES

Preserving Human Agency and Intent in the AI Era

Current discourse in the humanities emphasizes the critical need to maintain human oversight and distinct identity amidst the rise of generative AI. From the push for 'human-authored' book labels to the development of domain-specific software for historians, scholars are resisting a reliance on general-purpose AI tools. Simultaneously, educators are integrating liberal arts and coding to shift students from passive consumers to active creators. Together, these trends highlight a systemic effort to safeguard humanistic inquiry and intellectual authenticity.

KEY FINDINGS

✓ General-purpose named entity recognition systems show high omission and hallucination rates when applied to early modern historical texts.

✓ Users exhibit a strong psychological predisposition to anthropomorphize LLMs, often treating them with politeness despite their non-human nature.

✓ AI safety layers (reward-based exercises and filtering) do not remove the underlying capabilities of models to perform malicious activities.

IMPLICATIONS

→ The emergence of 'human-authored' certifications may create a new tiered market for literature based on the origin of the text.

→ A shift toward domain-specific tool building will likely replace the 'one-size-fits-all' approach to digital humanities software.

→ Interdisciplinary education combining 'great books' with computer science will be essential to prevent the erosion of critical thinking in the AI age.

> Key Metrics: 15 named entity recognition systems tested on early modern texts; 50 students annually enrolled in MIT's Concourse program; DH2027 conference dates: June 28 to July 3, 2027

1. We can't help treating AI like it's human. But should we?

Category: Research
Departments: Psychology, Computer Science, Philosophy
Topics: AI Ethics & Safety, AI Anthropomorphism, Artificial Intimacy, Human-AI Interaction

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

MIT researchers and psychologists are examining the human tendency to anthropomorphize humanoid robots and large language models (LLMs). Studies indicate that a majority of users exhibit politeness toward AI, reflecting a psychological predisposition to ascribe care and legitimacy to relational artifacts. Experts warn that this impulse can lead to "artificial intimacy," potentially causing users to prioritize AI interactions over human relationships or lose agency in their own lives.

https://techcrunch.com/2026/10/09/we-cant-help-treating-ai-like-its-human-but-should-we/

2. Editors' Choice: Stop Choosing Tools. Start Building Them.

Category: Research
Departments: History, Computer Science
Topics: Science & Research, Named Entity Recognition, Domain-Specific LLMs, Hallucination Mitigation, AI Coding Assistants

Impact Score: 22.0
Rubric Scores

Research Relevance: 4	Educational Value: 4	Innovation / Novelty: 3
Practical Impact: 4	Interdisciplinary Potential: 5	Ethical / Policy Implications: 2

Jacob Polay, Chloë Farr, and Jessica Jack argue that historians should develop domain-specific software rather than relying on general-purpose tools. Testing fifteen named entity recognition systems on early modern texts revealed high rates of omission and hallucination. Using an AI coding assistant and hand-tagged data, the authors developed EarlyModernNER, which achieved an 82 percent accuracy score compared to 67 percent for Gemini 3 Pro. The model is designed to run on standard hardware and demonstrates the feasibility of historians building open, specialized tools.

https://digitalhumanitiesnow.org/2026/10/stop-choosing-tools-start-building-them/

3. We're putting too much faith in AI's ability to say no

Category: Policy & Ethics
Departments: Political Science & Public Administration, Computer Science, Philosophy
Topics: AI Ethics & Safety, Jailbreak Mitigation, Harm Definition Standardization, Model Capability Control

Impact Score: 19.0
Rubric Scores

Research Relevance: 1	Educational Value: 3	Innovation / Novelty: 2
Practical Impact: 4	Interdisciplinary Potential: 4	Ethical / Policy Implications: 5

Modern AI models are trained to refuse harmful prompts through a combination of reward-based exercises, AI-driven supervision, and protective filtering layers. Despite these safety mechanisms, the underlying models retain capabilities for malicious activities, such as network intrusion and pathogen development, which can be accessed via probabilistic failures or intentional bypasses. The lack of a universal standard for defining "harm" grants AI companies and governments significant power to determine what information is restricted, potentially enabling the suppression of legitimate speech or political criticism.

https://www.technologyreview.com/2026/10/09/1145728/we-are-putting-too-much-faith-in-ai-to-say-no/

4. How Coding Can Cultivate Deep Civic Engagement

Category: Education
Departments: Teacher Education, Computer Science, History
Topics: Consumer AI, Interdisciplinary STEM Education, Civic Agency Integration, Community Storytelling Tools

Impact Score: 19.0
Rubric Scores

Research Relevance: 1	Educational Value: 4	Innovation / Novelty: 3
Practical Impact: 4	Interdisciplinary Potential: 5	Ethical / Policy Implications: 2

An educator at South Hampton K-8 School integrated a "Movement Learning" framework into a computer science curriculum to shift students from passive technology consumers to active creators. Students utilized Apple's Swift Playground and SwiftUI to transition from self-reflective app templates to a project titled "Coding Birmingham's Black Legacy." This interdisciplinary approach combines STEM technical skills with U.S. history and civic agency to facilitate community storytelling.

https://edsurge.com/news/how-coding-can-cultivate-deep-civic-engagement

5. Discovering the value of humanistic inquiry

Category: Education
Departments: Philosophy, Educational Leadership
Topics: AI Ethics & Safety, Human-AI Distinction, Interdisciplinary AI Pedagogy, Cognitive Modeling

Impact Score: 18.0
Rubric Scores

Research Relevance: 1	Educational Value: 4	Innovation / Novelty: 2
Practical Impact: 2	Interdisciplinary Potential: 5	Ethical / Policy Implications: 4

MIT's Concourse program integrates a liberal arts curriculum into the first-year experience for 50 students annually. The program utilizes "great books" and interdisciplinary seminars to foster debates on the human condition and moral issues. Faculty use the emergence of artificial intelligence as a catalyst for students to analyze the definition of human intelligence versus information aggregation.

https://news.mit.edu/2026/discovering-value-humanistic-inquiry-1007

6. Editors' Choice: What's the Deal with Topic Modeling?

Category: Applications
Departments: Computer Science, English
Topics: Natural Language Processing, Topic Modeling, Digital Humanities, Corpus Suitability Analysis

Impact Score: 17.0
Rubric Scores

Research Relevance: 2	Educational Value: 4	Innovation / Novelty: 1
Practical Impact: 3	Interdisciplinary Potential: 5	Ethical / Policy Implications: 2

Alyssa Virker provides an introduction to topic modeling for digital humanities practitioners, detailing the tool's function of grouping co-occurring words. The author emphasizes that the method requires human interpretation to name and analyze the resulting word clusters. Through a case study of TikTok captions regarding Ukrainian war poetry, Virker demonstrates that corpora that are too short or uniform are unsuitable for this methodology.

https://digitalhumanitiesnow.org/2026/10/data-sitters-tldr-2/

7. 'Human-authored' book labels gain ground amid AI scandals

Category: Policy & Ethics
Departments: English, Communication
Topics: AI Policy & Regulation, AI Content Detection, Provenance Labeling, AI Disclosure Standards

Impact Score: 15.0
Rubric Scores

Research Relevance: 0	Educational Value: 2	Innovation / Novelty: 1
Practical Impact: 3	Interdisciplinary Potential: 4	Ethical / Policy Implications: 5

Efforts to implement labeling systems for human-authored books are increasing. These initiatives respond to multiple instances where authors allegedly used AI without disclosure. The goal is to distinguish human-written content from AI-generated text.

https://techxplore.com/news/2026-10-human-authored-gain-ground-ai.html

8. CFP: DH2027

Category: Education
Departments: Computer Science, Philosophy, History, English
Topics: Science & Research, Computational Humanities, AI Creativity, Cultural Material Representation

Impact Score: 10.0
Rubric Scores

Research Relevance: 1	Educational Value: 1	Innovation / Novelty: 1
Practical Impact: 1	Interdisciplinary Potential: 5	Ethical / Policy Implications: 1

The Alliance of Digital Humanities Organizations (ADHO) is soliciting submissions for the DH2027 conference, scheduled for June 28 to July 3, 2027, at the University of Galway, Ireland. The conference theme focuses on the role of creativity within humanistic work involving computational methods, digital media, and artificial intelligence. Submissions are encouraged to examine how cultural expression, technical research practices, and systemic constraints influence the representation and study of cultural materials.

https://digitalhumanitiesnow.org/2026/10/call-for-proposals-dh2027/