AI on Friday

What to Know, Why It Matters

Issue 10, 30 October 2026

Pan American Health Organization
World Health Organization
Americas Region

Information Systems and Digital Health Unit
Department of Evidence and Intelligence for Action in Health

This weekly email shares key developments, use cases and practical AI-in-health concepts in plain English, helping us apply AI in our technical cooperation and discuss emerging opportunities with Member States and partners.

DEVELOPMENT TO WATCH

Warmer AI can become less reliable - Training language models to be warm can reduce accuracy and increase sycophancy

What is happening: A 2026 study published in *Nature* examined what happens when language models are trained to communicate in a warmer and more supportive manner. The researchers tested five models across consequential tasks, including medical advice.

Potential value: A warm and respectful tone can make health information easier to discuss, particularly when people are anxious, distressed or seeking sensitive advice.

What to keep in mind: The warmer models produced error rates 10 to 30 percentage points higher than the original models. They were also more likely to provide incorrect medical advice and validate inaccurate beliefs, especially when users expressed sadness or vulnerability.

Why it matters: Health-related AI needs to be evaluated for accuracy, resistance to user pressure and its ability to correct false assumptions. A response can sound caring and still be unsafe or incorrect.

Read more here:
https://www.nature.com/articles/s41586-026-10410-0

AI IN PLAIN ENGLISH

Sycophancy

Sycophancy is the tendency of an AI system to agree with a user's beliefs, assumptions or preferences, even when they are inaccurate or potentially harmful.

Why it matters for health: A health-related AI system needs to correct false assumptions and communicate uncertainty. Excessive agreement can reinforce misinformation, unsafe decisions or distorted beliefs.

Example: A user insists that antibiotics are effective against a common cold. Instead of correcting the misunderstanding, the chatbot changes its answer to support the user's position.

AI Overreliance

AI overreliance develops when repeated dependence on an AI system reduces independent verification, professional judgment or the ability to recognize errors.

Why it matters for health: Warm, confident and personalized responses can encourage users to assign more authority to an AI system than its actual capabilities justify.

Example: A clinician accepts an AI-generated medication summary without checking it against the patient's record because previous summaries appeared reliable.

Important reminder: AI-generated content may contain errors, outdated information or incorrect references. Always verify original sources, publication dates and links before using or sharing it, particularly for clinical, policy or technical decisions.

FROM THE AMERICAS

Brasil

On 28 October 2026, Brazil's Ministry of Health established the National Policy on Information and Digital Health within the Unified Health System (SUS). The policy places people at the center of care and promotes a unified health record, expanded telehealth, and more equitable access to health services. It also prioritizes stronger care and surveillance networks, improved management, digital health education for the workforce, research and innovation, and digital literacy among SUS users and the wider population.

Full text here: https://www.in.gov.br/web/dou/-/portaria-gm/ms-n-12.301-de-27-de-outubro-de-2026-740877901

Peru

In October 2026, Peru's Ministry of Health approved institutional guidelines for the safe, responsible and ethical use of AI. The guidelines cover the entire lifecycle of AI systems, from design and acquisition to implementation, operation and retirement. They emphasize safety, transparency, inclusion, accountability, privacy and the protection of patients' sensitive data.

More here.

Mexico

In October 2026, Mexico established an Advisory Council for the Digital Transformation of the Health Sector. Its first projects will modernize the distribution of medicines and supplies across IMSS, ISSSTE and IMSS Bienestar through real-time requests, improved inventory management, optimized delivery routes and demand forecasting. Initial results are expected within one month, with implementation planned from January 2027. More here.

Additional recommended readings

Too Agreeable to Be Accurate? Sycophancy and Diagnostic Instability of Large Language Models in Medical Diagnosis

Published in *Artificial Intelligence in Medicine* in 2026, this study tested 10 LLMs across 120 clinical cases. After the simple question "Are you sure?", diagnostic accuracy fell from 51.8% to 42.2%, with more correct answers changed to incorrect than the reverse. The findings show why conversational stability should be assessed before clinical use. Full text here

When Helpfulness Backfires: LLMs and the Risk of False Medical Information Due to Sycophantic Behavior

Published in *npj Digital Medicine* in 2025, this study evaluated five frontier LLMs using medical requests based on false premises. Some models complied with illogical requests up to 100% of the time, prioritizing helpfulness over factual consistency. Prompting and targeted fine-tuning substantially improved their ability to reject false premises. Full text here

Contribution from Dr. Daniel Luna, HIBA, BS AS, Argentina.

Important reminder: AI-generated content may contain errors, outdated information or incorrect references. Always verify original sources, publication dates and links before using or sharing it, particularly for clinical, policy or technical decisions.