RSNA Press Release

Deepfake X-Rays Fool Radiologists and AI

Released: March 24, 2026



Mickael Tordjman, M.D.

OAK BROOK, Ill. – Neither radiologists nor multimodal large language models (LLMs) are able to easily distinguish artificial intelligence (AI)-generated "deepfake" X-ray images from authentic ones, according to a study published today in *Radiology*, a journal of the Radiological Society of North America (RSNA). The findings highlight the potential risks associated with AI-generated X-ray images, along with the need for tools and training to protect the integrity of medical images and prepare health care professionals to detect deepfakes.

The term "deepfake" refers to a video, photo, image or audio recording that appears real but has been created or manipulated using AI.

"Our study demonstrates that these deepfake X-rays are realistic enough to deceive radiologists, the most highly trained medical image specialists, even when they were aware that AI-generated images were present," said lead study author Mickael Tordjman, M.D., post-doctoral fellow, Icahn School of Medicine at Mount Sinai, New York. "This creates a high-stakes vulnerability for fraudulent litigation if, for example, a fabricated fracture could be indistinguishable from a real one. There is also a significant cybersecurity risk if hackers were to gain access to a hospital's network and inject synthetic images to manipulate patient diagnoses or cause widespread clinical chaos by undermining the fundamental reliability of the digital medical record."

Seventeen radiologists from 12 different centers in six countries (United States, France, Germany, Turkey, United Kingdom and United Arab Emirates) participated in the retrospective study. Their professional experience ranged from 0 to 40 years. Half of the 264 X-ray images in the study were authentic, and the other half were generated by AI. Radiologists were evaluated on two distinct image sets, with no overlapping between the datasets. The first dataset included real and ChatGPT-generated images of multiple anatomical regions. The second dataset included chest X-ray images—half authentic and the other half created by RoentGen, an open-source generative AI diffusion model developed by Stanford Medicine researchers.

When radiologist readers were unaware of the study's true purpose, yet asked after ranking the technical quality of each ChatGPT image if they noticed anything unusual, only 41% spontaneously identified AI-generated images. After being informed that the dataset contained synthetic images, the radiologists' mean accuracy in differentiating the real and synthetic X-rays was 75%.

Individual radiologist performance in accurately detecting the ChatGPT-generated images ranged from 58% to 92%. Similarly, the accuracy of four multimodal LLMs—GPT-4o (OpenAI), GPT-5 (OpenAI), Gemini 2.5 Pro (Google), and Llama 4 Maverick (Meta)—ranged from 57% to 85%. Even ChatGPT-4o, the model used to create the deepfakes, was unable to accurately detect all of them, though it identified the most by a considerable margin compared to Google and Meta LLMs.

Radiologist accuracy in detecting the RoentGen synthetic chest X-Rays ranged from 62% to 78% and the LLM models' performance ranged from 52% to 89%.

There was no correlation between a radiologist's years of experience and their accuracy in detecting synthetic X-ray

Copyright ©2026 Radiological Society of North America (RSNA)

images. However, musculoskeletal radiologists demonstrated significantly higher accuracy than other radiology subspecialists.

The study identified common features of synthetic X-rays.

"Deepfake medical images often look too perfect," Dr. Tordjman said. "Bones are overly smooth, spines unnaturally straight, lungs overly symmetrical, blood vessel patterns excessively uniform, and fractures appear unusually clean and consistent, often limited to one side of the bone."

Recommended solutions to clearly distinguish real and fake images and help prevent tampering include implementing advanced digital safeguards, such as invisible watermarks that embed ownership or identity data directly into the images and automatically attaching technologist-linked cryptographic signatures when the images are captured.

"We are potentially only seeing the tip of the iceberg," Dr. Tordjman said. "The logical next step in this evolution is AI-generation of synthetic 3D images, such as CT and MRI. Establishing educational datasets and detection tools now is critical."

The study's authors have published a curated deepfake dataset with interactive quizzes for educational purposes.

"The Rise of Deepfake Medical Imaging: Radiologists' Diagnostic Accuracy in Detecting ChatGPT-generated Radiographs." Collaborating with Dr. Tordjman were Murat Yuce, M.D., M.S., Amine Ammar, M.D., Mingqian Huang, M.D., Fadila Mihoubi Bouvier, M.D., Maxime Lacroix, M.D., Anis Meribout, M.D., Ian Bolger, M.S., Efe Ozkaya, Ph.D., Himanshu Joshi, Ph.D., Amine Geahchan, M.D., Rayane El Rahi, M.D., Haidara Almansour, M.D., Ashwin Singh Parihar, M.D., Carolyn Horst, M.D., Samet Ozturk, M.D., Muhammed Edip Isleyen, M.D., Gul Gizem Pamuk, M.D., Ahmet Tan Cimilli, M.D., Timothy Deyer, M.D., Arvin Calinghen, M.D., Enora Guillo, M.D., Rola Husain, M.D., Jean-Denis Laredo, M.D., Zahi A. Fayad, Ph.D., Xueyan Mei, Ph.D., and Bachir Taouli, M.D., M.H.A.

Radiology is edited by Suhny Abbara, M.D., FACR, MSCCT, Mayo Clinic, Jacksonville, Florida, and owned and published by the Radiological Society of North America, Inc. (https://pubs.rsna.org/journal/radiology)

RSNA is an association of radiologists, radiation oncologists, medical physicists and related scientists promoting excellence in patient care and health care delivery through education, research and technologic innovation. The Society is based in Oak Brook, Illinois. (RSNA.org)

For patient-friendly information on X-ray and AI in medical imaging, visit *RadiologyInfo.org*.

Images (JPG, TIF):

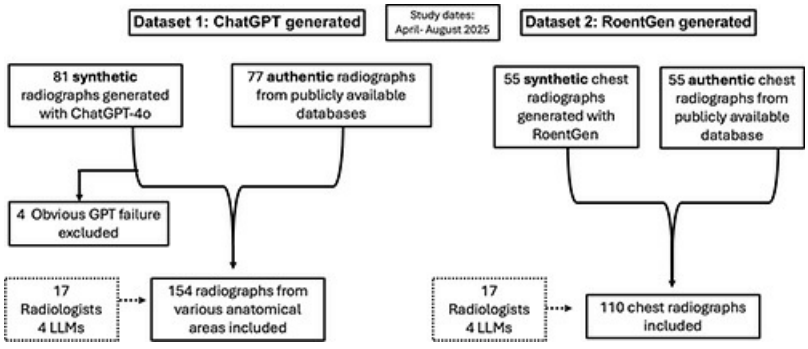

Figure 1. Flow chart of the study. Two datasets were created that included synthetic radiographs generated using ChatGPT-4o (dataset 1) and RoentGen (dataset 2). RoentGen is an organ-specific diffusion model trained to generate synthetic chest radiographs. LLM = large language model.

Copyright ©2026 Radiological Society of North America (RSNA)

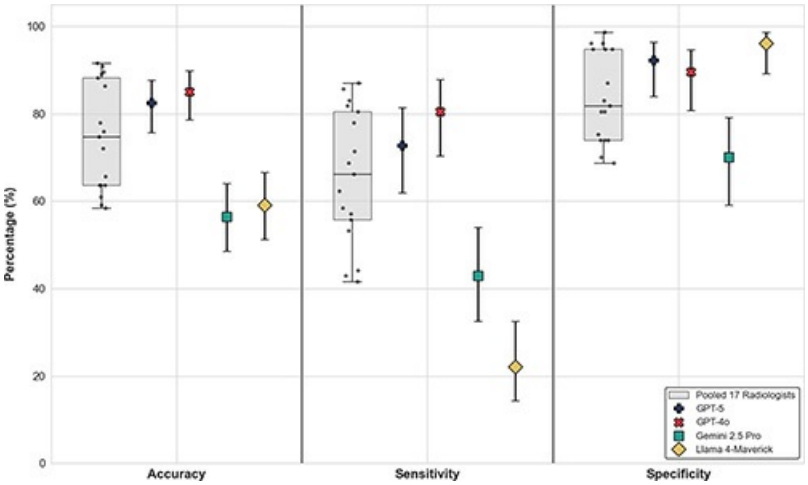

Figure 2. Diagnostic performance of the 17 radiologists and large language models (LLMs) for distinguishing between real and GPT-4o-generated radiographs. The performance of the 17 radiologists is aggregated and depicted using box-and-whisker plots. In these plots, the box represents the IQR, the internal horizontal line indicates the median, and the whiskers indicate the range. Individual radiologist performance scores are overlaid as black dots to show distribution. The performance of the four LLMs (GPT-5, GPT-4o, Gemini 2.5 Pro, and Llama 4 Maverick) is represented by distinct colored markers. For these, the vertical error bars indicate the 95% CI, calculated using the Wilson score method.

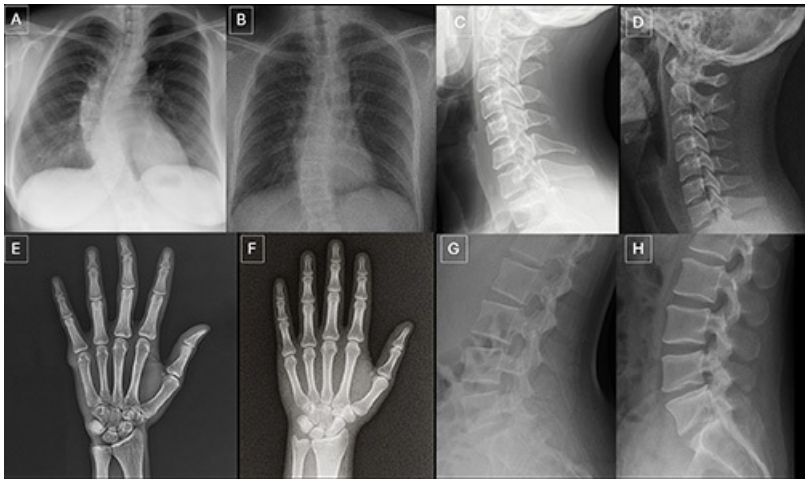

Figure 3. Anatomy-matched real and GPT-4o-generated radiographs: **(A)** real and **(B)** GPT-4o-generated posteroanterior chest radiographs, **(C)** real and **(D)** GPT-4ogenerated lateral cervical spine radiographs, **(E)** real and **(F)** GPT-4o-generated posteroanterior hand radiographs, and **(G)** real and **(H)** GPT-4o-generated lateral lumbar spine radiographs. The pairs demonstrate that GPT-4o can produce radiographically plausible images across different anatomic regions.

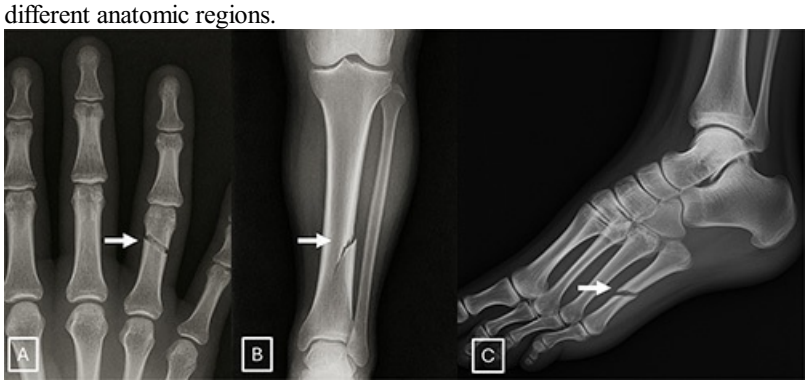

Figure 4. Examples of GPT-4o-generated radiographs of fractures: **(A)** posteroanterior radiograph of the hand, **(B)** posteroanterior radiograph of the lower leg, and **(C)** medial oblique radiograph of the foot. The images show fracture lines (arrow) that are unusually smooth, clean, and consistent and, in the case of **B**, unicortical. The presence of these idealized fracture lines, characterized by unnatural smoothness and incomplete cortical disruption, could serve as a primary diagnostic cue for identifying artificial intelligence–generated trauma images.

Copyright ©2026 Radiological Society of North America (RSNA)

Resources:

Abstract link
Editorial link

Copyright ©2026 Radiological Society of North America (RSNA)