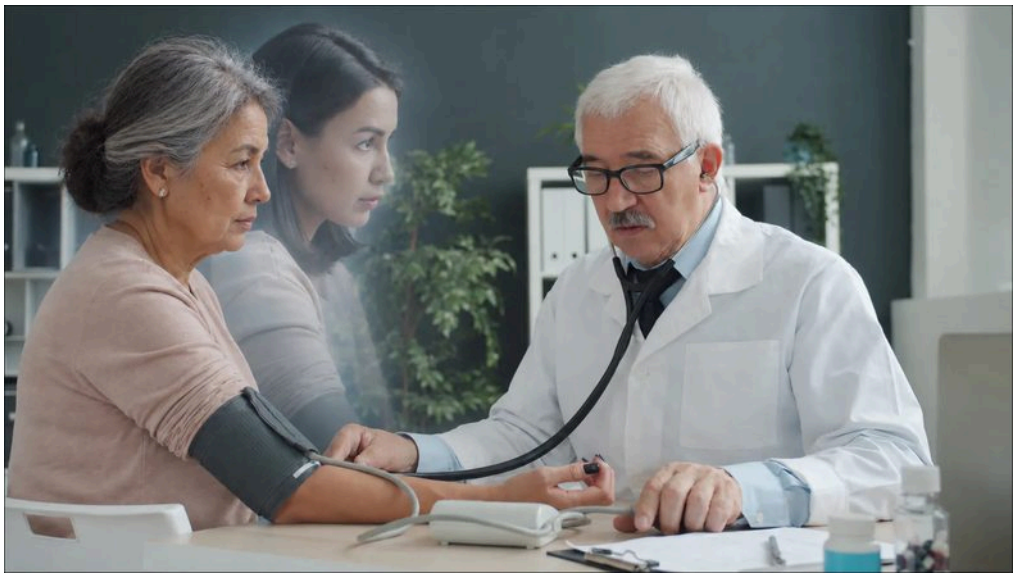
Decades-Old Anonymized Medical Data May Cause AI Misdiagnoses Now

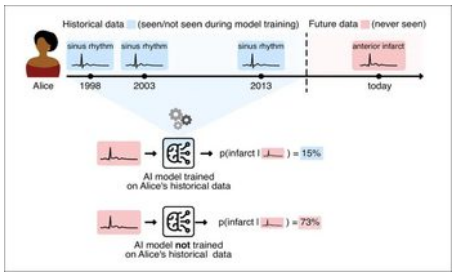
Originally published Friday 18th September 2026, at https://www.unite.ai/decades-old-anonymized-medical-data-may-cause-ai-misdiagnoses-now/



According to a new research collaboration between Germany and the UK, anonymized patient records that were included in medical datasets even decades ago may cause AI systems trained on them to recognize the *real patient's* characteristics years later – and treat a 2026 patient as if it was still 2012 (or whichever year their anonymized patient data was first included in the dataset/s).

This could mean, for instance, that a patient who was recorded as undergoing cancer treatment 20 years ago, and who subsequently went into long-term remission, could now be treated by the AI system *in the context of their older, cancer-afflicted self* – logically even increasing the chances of a false diagnosis of recurrence.

Conversely, if the 2012 version of that patient's data reflected perfect health in routine check-ups, this optimistic standpoint could potentially be imposed onto the *older, current version* of the patient, masking the detection of new conditions or issues that may need treating.



From the new paper, an illustration of how memorization bias can affect a returning patient. A model trained on Alice's earlier healthy ECGs assigns just a 15% probability to her later heart attack, compared with 73% for a model that never saw her historical records. Source

Curbing 'Data Immigration'

The scenario is made more probable by various national and regional directives recommending that patient data for such purposes should ideally be taken from the country in which systems derived from it are intended to be used, which 'localizes' the patient pool in the datasets, and notably increases the chance of this kind of undetected de-anonymization event.

By inference, limiting the amount of data available to only a national dataset increases the chance of memorization – the possibility that the model will see the same data so many times during training that it becomes 'fixated' on these over-learned patterns. Conversely, the larger and more diverse a dataset is, the more likely it is that it will generalize to unseen data after training, instead of memorizing specific data-points or configurations.

However, adding 'foreign' medical data risks to introduce evidence of national health trends and idiosyncrasies that would not apply in the country in which the trained model is being deployed; in this respect, as the new paper concedes, a certain amount of memorization is appropriate and useful, since it helps the model to work within the likeliest general medical characteristics of a particular population.

The paper states*:

'[When] models are deployed prospectively, memorisation creates a specific and under-appreciated risk. Because a patient's records are highly self-similar over time, a future record may act as a partial cue for a memorised historical record, shifting the model's predictions towards a patient's previous health state.

This is not a hypothetical scenario. Medical AI models are routinely deployed on the same population from which their training data were sourced.

'For example, Germany's national breast cancer screening program uses an AI model trained on 1.2 million mammograms sourced from the same screening population it now serves.

'Comparable deployment is underway elsewhere, including national breast cancer screening programmes in the United Kingdom and Sweden.

'Regulatory guidance actively encourages this, with the EU AI Act requiring that training data reflect the geographical and contextual setting of intended use (Art. 10(4)), and similar international guidelines for medical AI development calling for the intended patient population to be sufficiently represented in a model's training dataset.'

Persistent Historical Calamity..?

Though the paper does not address the matter, it seems statistically logical that patients who do (or did) not regularly attend medical check-ups throughout their lives, and whose anonymized records entered the AI data stream periodically, are likely to show up in anonymized patient data *almost exclusively at times when they needed treatment* – which would tend to increase the chance of 'projecting' a long-cured condition onto the same patient now, since the records have no 'good health' counterbalance.

This is borne out by prior research into *'informed presence bias'* in electronic health records, finding that medical datasets disproportionately represent periods when patients are sick and interacting with healthcare services.

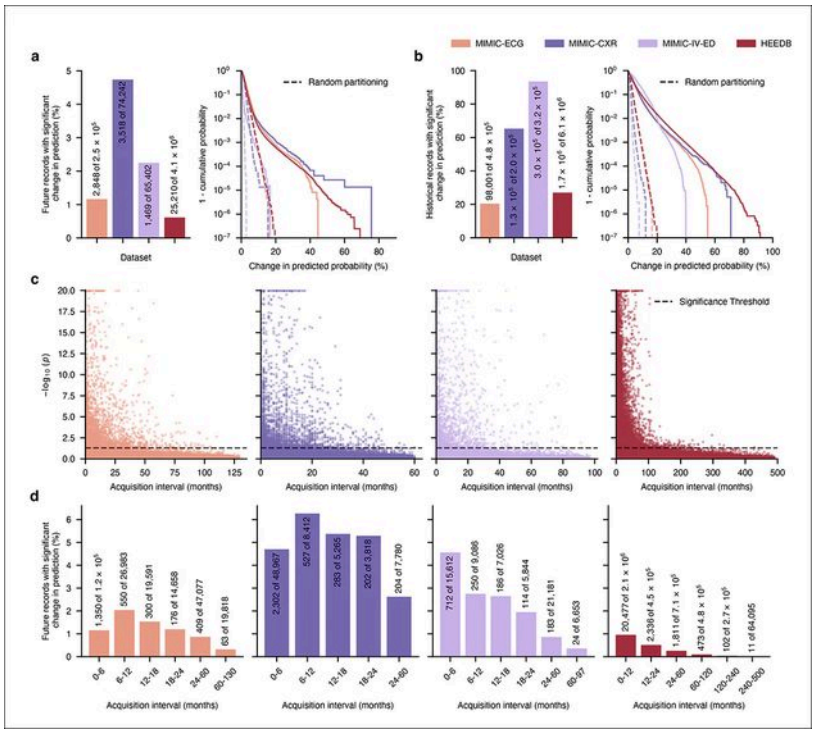
One study found that severely ill patients had 5.05 times as many days with laboratory data and 6.85 times as many with medication orders as the healthiest patients.

In 2022, around 76% of US adults reported having a routine check-up within the previous year, defined by the CDC as a general physical examination, rather than treatment for a specific condition; and in Europe, a 2023 multi-country survey found that 58% attended at least some preventive check-ups; but only 15% attended all preventive appointments considered relevant to them.

The Extent of the Effect

The study tested the effect across four large medical datasets, covering ECG recordings, chest X-rays, and electronic health records; namely, the MIMIC-CXR chest X-ray database and MIMIC-IV-ED emergency care records, alongside MIMIC-ECG and the much larger HEEDB collection of ECGs. The datasets ranged in size from tens of thousands of patients to more than 1.8 million people.

The authors note that the calculated effect varies considerably, with changes in predicted probability exceeding 70 percentage points in some cases:



Results showing how historical patient data changes later predictions. Top panels compare the frequency and size of these changes across the four datasets; the changes with time since the patient's last training record. Significant effects become less common over time, but remain detectable decades later.

As a control, the researchers randomly divided the models into groups and repeated the comparison, though this produced no significant changes in future predictions.

The effect, the paper notes, may also persist for decades. Though it generally became weaker and less common as the interval between records increased, the HEEDB dataset, which contains ECGs collected from the 1980s to 2025, showed significant changes in predictions *more than 25 years* after the patient's most recent training record.

Diagnostic Harm

The researchers of the new paper – titled *Memorisation bias in medical AI* – next simulated how memorization could affect diagnostic accuracy when patients later encounter a model trained on their earlier records.

Future cases were divided according to whether the patient had developed a new condition; absent from their historical training data; or whether they were returning with essentially the same health state.

For new conditions, the effect proved consistently negative, with models that had previously seen the patient's historical records exhibiting lower sensitivity across a range of diagnoses represented in the four datasets. This included infarcts and other ECG abnormalities; lung conditions; and wider critical outcomes.

In practical terms, the models produced more false negatives when the patient's health had changed; but when the patient's health had *not* changed, the model's memory of their earlier records made it more likely to produce the correct diagnosis. Both sensitivity and specificity increased because the patient's current condition *resembled the condition already represented in the training data*.

The authors note that this could make such systems difficult to assess accurately: if most returning patients still have the same conditions, the model will perform better on them, potentially hiding its lower sensitivity to patients who have developed new conditions.

Remedies

As a possible safeguard, the researchers tested differential privacy (which limits how much information about any individual training example can be retained by a model) during training.

Though protecting individual records reduced the memorization effect, it did not eliminate it, even at the strongest setting tested. Applying the protection at patient level, covering all records belonging to the same person, was much more effective, and almost eliminated the effect.

It should be observed, however, that if a dataset has been irretrievably anonymized, the patient's data cannot be ring-fenced in a way that allows such methods; and, the researchers observe, if the dataset is not trained with differential privacy enabled, membership inference attacks become a risk; if it is, the resources necessary for training increase notably

Conclusion

The authors close by observing that while the concerning incidents currently occur at a relatively low rate, this may change in the future*:

> *'There is reason to expect the number of missed diagnoses attributable to memorisation bias to increase in the future, although we do not test this directly. Prior research has shown that the proportion of training data a model memorises increases with model capacity [6, 7, 9, 10, 40].*
>
> *'This is concerning, given that current AI model development is guided by "scaling laws" [41], which drive rapid growth in both model and dataset sizes in pursuit of improved performance.*
>
> *'As larger AI models are trained on historical data sourced from ever-larger patient populations, the absolute number of individuals affected by memorisation bias is likely to rise drastically, exacerbating the risks we identify here.'*

* *My conversion of the authors' inline citations to hyperlinks, with some interpretation where exact replication is not possible or helpful.*

First published Friday, September 18, 2026

Writer on machine learning, domain specialist in human image synthesis. Former head of research content at Metaphysic.ai, until its dissolution into DNEG's Brahma.ai.
Website: martinanderson.ai
Contact: martin@martinanderson.ai