Future Healthcare Journal 2022 Vol 9, No 2: 190–3

DIGITAL TECHNOLOGY

Generative adversarial networks and synthetic patient data: current challenges and future perspectives

Authors: Anmol Arora[A] and Ananya Arora[A]

ABSTRACT

Artificial intelligence (AI) has been heralded as one of the key technological innovations of the 21st century. Within healthcare, much attention has been placed upon the ability of deductive AI systems to analyse large datasets to find patterns that would be unfeasible to program. Generative AI, including generative adversarial networks, are a newer type of machine learning that functions to create fake data after learning the properties of real data. Artificially generated patient data has the potential to revolutionise clinical research and protect patient privacy. Using novel techniques, it is increasingly possible to fully anonymise datasets to the point where no datapoint is traceable to any real individual. This can be used to expand and balance datasets as well as to replace the use of real patient data in certain contexts. This paper focuses upon three key uses of synthetic data: clinical research, data privacy and medical education. We also highlight ethical and practical concerns that require consideration.

KEYWORDS: machine learning, generative adversarial networks, confidentiality, ethics, legal frameworks

DOI: 10.7861/fhj.2022-0013

Introduction

Artificial intelligence (AI) has the potential to transform healthcare, '*by helping clinicians to make more accurate diagnoses, providing decision support, and automating tasks. AI has already been used to develop predictive models for conditions such as heart disease, cancer, and diabetes. These models can help clinicians to identify patients who are at risk of developing a particular condition and to recommend treatments.*'[1]

This introductory quote was written entirely by a generative AI system from the prompt 'Artificial intelligence (AI) has the potential to transform healthcare'. The system in question is the GPT-3 model, developed by OpenAI (San Francisco, USA).[1] Similar generative AI systems have been able to produce extremely realistic pictures of human faces and, in a medical context, synthetic chest X-rays that are indistinguishable from real ones. Broadly, there are two main types of AI: deductive and generative. Research into uses of AI so far has mostly focused on deductive rather than generative systems. Deductive algorithms are increasingly capable of analysing data to find patterns that would be unfeasible for humans to program; they may be used in data analysis and even diagnosis. However, generative AI systems have received much less attention than their deductive counterparts from the media and the broader research community. In the context of healthcare, they provide potential opportunity to create training material in the form of text, video, audio and even simulations.[2] The fake datasets are constructed such that they retain all the properties and patterns of the original data, without being attributable to real individuals. As a result, data can be shared without traditional data concerns since the data being shared is artificial. It is important to recognise that, despite the overly optimistic introduction to AI in healthcare by the GPT-3 model earlier, both deductive and generative models are yet to demonstrate a real-world impact in improving patient outcomes. In particular, there is still a shortage of randomised controlled trials.

In the UK, the use of patient data is governed by the Caldicott principles, which dictate that confidential data should be used minimally and only when necessary.[3] As synthetic data begins to match the clinical utility of real data, it is difficult to envisage a future in which synthetic data is not given more credence than real data for clinical research.

Generative adversarial networks

One type of generative AI is 'generative adversarial networks' (GANs), which uses two competing AI models to produce synthetic data (shown in Fig 1). This has been widely applied to create fake images or 'deepfakes'. The system consists of two machine learning algorithms: the generator and the discriminator. The generator aims to create fake images, starting with producing random noise and progressively producing more realistic data. The discriminator aims to determine whether the output of the generator is real or fake. For each image produced by the generator, the discriminator produces a binary classification of real or fake. Initially, it would be very easy for the discriminator because it is comparing real images with random noise. However, over time the generator learns from the output

Authors: [A]medical student, University of Cambridge, Cambridge, UK

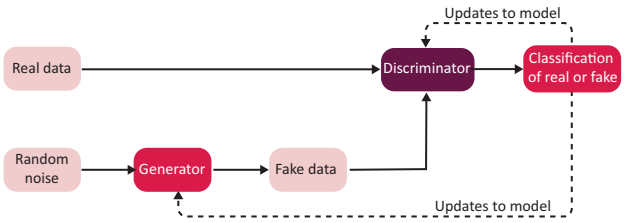

Fig 1. The generative adversarial network. A generator produces synthetic data that is judged by the discriminator as appearing real or fake.

of the discriminator in a feedback loop. Eventually the output of the generator would be so realistic that the discriminator is unable to determine whether the fake images are real or fake. The success rate of classification would approximate to 50%, effectively guesswork.

Producing synthetic data is a highly resource intensive process, especially for complex images such as medical data. Despite recent advances, generating data remains computationally intensive and it can take weeks to develop successful models.[4,5] It also requires access to large data storage capacity and powerful graphics processing units (GPUs). Fig 2 shows a generation of fake chest X-rays using NVIDIA's StyleGAN2-ADA model.[4] Over time, the output of the GAN improves from producing random noise to realistic X-rays. In recent years, GANs have also been used to produce synthetic computed tomography (CT), magnetic resonance imaging (MRI) and positron emission tomography as well as retinal, dermascopic and ultrasound images.[6]

Herein, we outline three potential uses of synthetic data in healthcare:

- clinical research
- medical education
- protecting patient privacy.

Clinical research

In clinical research, generative AI may be used to create synthetic data to enhance datasets and increase diversity. Perhaps the simplest use is the ability to increase the size of datasets for research. This is particularly important when those datasets are being used to train other machine learning algorithms since the success of such an algorithm is a function of the sample size of the input data. Where datasets are imbalanced and not representative of the population they aim to serve, generative AI in the form of synthetic minority oversampling technique (SMOTE) may be used to selectively augment the representation of minority datapoints.[7] In this way, there is potential for synthetic data to help mitigate algorithmic bias in healthcare uses of machine learning, both in constructing algorithms and responding to dataset shift.[8,9] Synthetic data may also be used to audit medical applications of machine learning by exposing algorithms to novel simulated data in adversarial testing.[10]

Similarly, GANs may also be used in the construction of synthetic digital twins, whereby an artificial construction of a system is created that preserves all the patterns of the true system.[11] If the synthetic data preserves the relationships, patterns and characteristics of the original data, then this may be used for subsequent analyses in place of real data, accelerating data acquisition, labelling and analysis. GANs may also be used in the domain of image-to-image translation, such as synthesising CT images from an MRI image.[12]

In clinical trials, there has been interest in the development of synthetic control arms, whereby placebo groups can be modelled based on historical information.[13] In cases where a synthetic control arm is suitable, a reduced need for a real-life placebo group can save costs and facilitate increased sizes of treatment arms of clinical trials; for example, if a trial is recruiting 10,000 participants in a treatment arm and 10,000 participants in a control arm, a novel trial methodology could instead utilise 15,000 in the treatment arm, 5,000 in the control arm with 10,000 synthetic cases supplementing that control group. This would be particularly useful when trials are performed on patient groups with limited overall population sizes.

Medical education

In medical education, generative AI may be used to rapidly produce training material and simulations for students to use for learning. Importantly, this training material is customisable; for example, if a student was having difficulty distinguishing between left lower lobe collapse and consolidation, examples of each type could be created and presented to the student. In this way, synthetic data can assist in presenting students with a higher proportion of 'edge-case' learning material and reducing the proportion of material presented that the student is already comfortable with.[14] Furthermore, since each image is unique, they cannot be reverse-searched during examinations and they are

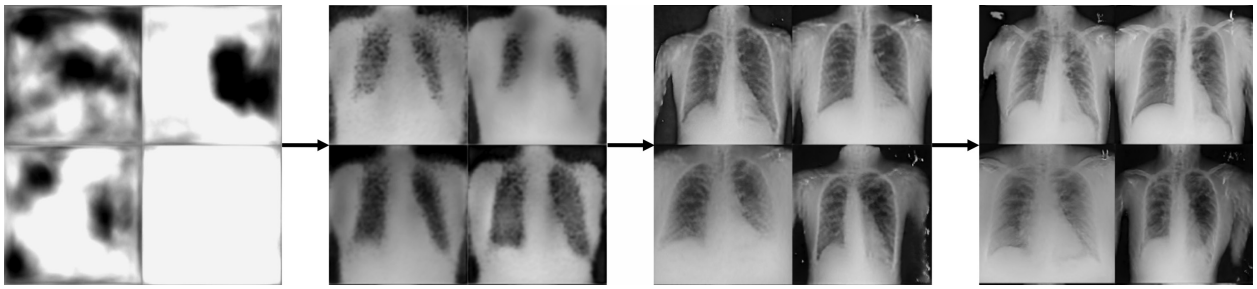

Fig 2. Progressive outputs of a generative artificial intelligence model to produce fake chest X-rays. The training time required to produce the right-most X-rays was approximately 14 hours; on close inspection, these are recognisable as being fake.

likely to differ from the examples that students have been shown in lectures and during revision. It has even been suggested that generative AI could be used to create virtual instructors, with AI generated characters.[15] This could be used to create 'fake' virtual patients with synthetic clinical presentations in the future.

Protecting patient privacy

Perhaps the most important use of synthetic data is to protect patient privacy and increase the scope for institutions to share data openly. There is an increasing threat of deanonymisation of patient data where identification of individuals is already possible from electroencephalography (EEG) and an increasing amount of personal information is being extractable from data such as retinal fundus photography and electrocardiography (ECG).[16–18] Such extractable personal information includes details of age, sex and disease status. There is a growing risk that data, which is currently thought to be anonymised, may be deanonymised in the future. Since an increasing amount of data is currently being shared, including in online repositories, there is an impending danger that this data may be maliciously deanonymised.

Synthetic data acts as a method of anonymisation that can be used to share data between institutions. When successful, no single synthetic data point is attributable to any individual patient but the overall patterns within the dataset can be maintained. Since no real patient data is being shared, the data can be shared freely. This may involve replacing open access datasets with synthetic datasets that contain no patient-attributable information but preserve all the patterns from the original data. As the data does not relate to a natural person or any individual, it has been suggested that the data is neither sensitive nor confidential.[19] Beyond healthcare, the utility of GANs for anonymisation has been reported in the field of finance, which also features regulatory and privacy requirements.[20]

Emerging limitations

There are two main costs associated with running GANs: computing power and labelling of synthetic data. Training GANs is computationally expensive and tends to require access to GPUs. These are increasingly accessible to a general audience through cloud computing services, such as Google Cloud and Amazon Web Services. Such services can be expensive and also carry an environmental burden due to the energy expenditure. In theory, once a GAN is trained, it can generate unlimited amounts of synthetic data. However, in the context of healthcare, this data is most useful if it is labelled. A well-recognised limitation of AI use in healthcare is the need for accurate ground-truth labelling. Failing to accurately label data that is then used for training other machine learning models may hinder their performance. In the future, a potential solution may be to label synthetic data with mature machine learning models trained on real data; however, human labelling would more likely be required in the foreseeable future.

There are ethical issues associated with GANs, including the risk that GANs may be used maliciously in a clinical context in such a way as to present synthetic data as real data. Other ethical concerns include the risk of synthetic data being used as a mechanism of evading data privacy laws. As synthetic data does not relate to an individual, it is not covered by standard data protection legislation and can be used to legally disseminate information beyond the borders of an organisation. While this information is still anonymous at an individual level, this could be problematic in scenarios where aggregate trends could potentially be misused; for example, in determining insurance premiums.[21] A related issue is that since the data are now anonymous, it could be subject to Freedom of Information requests if held by a public authority.

While there are metrics for assessing the fidelity of synthetic data, there is a need for recognised quantitative metrics that can measure the fidelity and anonymity of synthetic data compared with real data. The synthetic data must be sufficiently realistic enough that it can be used in place of genuine data, but it must also be sufficiently different enough so that original datapoints cannot be identified in order to prevent information leakage.[22] Reporting guidelines, similar to those for deductive models, may be required.[23]

Conclusion

Overall, GANs provide a great opportunity to enhance medical education and research, with the ultimate objective of improving patient care. In the context of research, these models are becoming increasingly capable at writing text that would be near-impossible to distinguish from a human's writing. They may soon become capable of writing their own full-length research papers, especially literature reviews, at which point, it must be asked how much of a right would the human using the system have claim to authorship.

The generative AI system that introduced this article, had an earlier full draft of this article as prompt and created this conclusion:

> *Overall, GANs offer great potential for enhancing medical education and research. However, there are some concerns that need to be addressed, such as the authorship of research papers and the creation of fake medical images and videos. Nonetheless, these concerns should not overshadow the great potential that GANs have to improve patient care.*[1] ■

Conflicts of interest

Anmol Arora has roles with the National Institute for Health Research (NIHR), Health Data Research UK, NHS England, NHS Improvement and Moorfield's Eye Hospital.

References

1 Brown TB, Mann B, Ryder N *et al*. Language models are few-shot learners. *arXiv* 2020.
2 Arora A. Disrupting clinical education: Using artificial intelligence to create training material. *Clin Teach* 2020;17:357–9.
3 The UK Caldicott Guardian Council. *The Caldicott Principles*. UKCGC. www.ukcgc.uk/the-caldicott-principles [Accessed 07 February 2022].
4 Karras T, Hellsten J. *NVlabs/stylegan2-ada-pytorch*. GitHub, 2022. https://github.com/NVlabs/stylegan2-ada-pytorch [Accessed 28 January 2022].
5 Peters D. *The promise and pitfalls of synthetic data*. University Affairs Affaires universitaires, 2021. www.universityaffairs.ca/news/news-article/the-promise-and-pitfalls-of-synthetic-data [Accessed 28 January 2022].
6 Jeong JJ, Tariq A, Adejumo T *et al*. Systematic review of generative adversarial Networks (GANs) for medical image classification and segmentation. *J Digit Imaging* 2022;35:137–52.
7 Elreedy D, Atiya AF. A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance. *Information Sciences* 2019;505:32–64.

8 Norori N, Hu Q, Aellen FM, Faraci FD, Tzovara A. Addressing bias in big data and AI for health care: A call for open science. *Patterns (N Y)* 2021;2:100347.
9 Finlayson SG, Subbaswamy A, Singh K *et al*. The clinician and dataset shift in artificial intelligence. *N Engl J Med* 2021;385: 283–6.
10 Liu X, Glocker B, McCradden MM *et al*. The medical algorithmic audit. *Lancet Digital Health* 2022;4:e384–97.
11 Xing X, Del Ser J, Wu Y *et al*. HDL: Hybrid deep learning for the synthesis of myocardial velocity maps in digital twins for cardiac analysis. *IEEE Journal of Biomedical and Health Informatics* 2022.
12 Wang C, Yang G, Papanastasiou G *et al*. DiCyc: GAN-based deformation invariant cross-domain information fusion for medical image synthesis. *Information Fusion* 2021;67:147–60.
13 Anju Life Sciences Software. *What clinical trial leaders need to know about using synthetic data*. Anju, 2020. www.anjusoftware.com/about/all-news/using-synthetic-data [Accessed 28 January 2022].
14 Arora A. Artificial intelligence: a new frontier for anaesthesiology training. *British Journal of Anaesthesia* 2020;125:e407–8.
15 Pataranutaporn P, Danry V, Leong J *et al*. AI-generated characters for supporting personalized learning and well-being. *Nat Mach Intell* 2021;3:1013–22.
16 Marcel S, Millán JDR. Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation. *IEEE Trans Pattern Anal Mach Intell* 2007;29:743–52.
17 Topol EJ. What's lurking in your electrocardiogram? *Lancet* 2021;397:785.
18 Poplin R, Varadarajan AV, Blumer K *et al*. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nature Biomedical Engineering* 2018;2:158–64.
19 WireWheel. *What privacy officers need to know about synthetic data*. WireWheel, 2021. https://wirewheel.io/privacy-synthetic-data [Accessed 28 January 2022].
20 Assefa S. Generating synthetic data in finance: opportunities, challenges and pitfalls. *Social Science Research Network* 2020. https://papers.ssrn.com/abstract=3634235 [Accessed 13 May 2022].
21 Arora A, Arora A. Synthetic patient data in health care: a widening legal loophole. *Lancet* 2022;399:1601–2.
22 Chen RJ, Lu MY, Chen TY, Williamson DFK, Mahmood F. Synthetic data in machine learning for medicine and healthcare. *Nat Biomed Eng* 2021;5:493–7.
23 Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med* 2020;26:1364–74.

Address for correspondence: Mr Anmol Arora, School of Clinical Medicine, University of Cambridge, Addenbrooke's Hospital, Hills Road, Cambridge CB2 0SP, UK.
Email: aa957@cam.ac.uk
Twitter: @AnmolArora_98