Generative AI models in time varying biomedical data: a systematic review

Rosemary Yuan He, Varuni Sarwal, Xinru Qiu, Yongwen Zhuang, Le Zhang, Yue Liu, Jeffery N Chiang

Submitted to: Journal of Medical Internet Research
on: April 30, 2024

Table of Contents

Generative AI models in time varying biomedical data: a systematic review

Rosemary Yuan He[1, 2] BSc, MCS; Varuni Sarwal[1, 2] MCS; Xinru Qiu[3] PHD; Yongwen Zhuang[4] PHD; Le Zhang[5] MCS; Yue Liu[6] MCS; Jeffery N Chiang[1, 7] PHD

[1]Department of Computational Medicine University of California, Los Angeles Los Angeles US
[2]Department of Computer Science University of California, Los Angeles Los Angeles US
[3]Division of Biomedical Sciences School of Medicine University of California Riverside Riverside US
[4]Department of Biostatistics University of Michigan Ann Arbor US
[5]Institute for Integrative Genome Biology University of California Riverside Riverside US
[6]Institute for Cellular and Molecular Biology University of Texas at Austin Austin US
[7]Departments of Neurosurgery David Geffen School of Medicine UCLA Los Angeles US

Corresponding Author:
Jeffery N Chiang PHD
Departments of Neurosurgery
David Geffen School of Medicine
UCLA
300 Stein Plaza, Suite 560
Los Angeles
US

Abstract

Background: Trajectory modeling is a longstanding challenge in the application of computational methods to healthcare. However, traditional statistical and machine learning methods do not achieve satisfactory results as they often fail to capture the complex underlying distributions of multi-modal health data, and long-term dependencies throughout patients' medical histories. Recent advances in generative AI have provided powerful tools to represent complex distributions and patterns with minimal underlying assumptions. These have had a major impact in fields such as finance and environmental sciences, and recently researchers have turned to these methods for disease modeling.

Objective: While AI methods have proven powerful, their application in clinical practice remains limited due to their black-box like nature. The proliferation of AI algorithms poses a significant challenge for non-developers to track and incorporate these advances into clinical research and application. In this work, we survey peer-reviewed, generative AI model papers with specific applications in time series health data.

Methods: Our search includes single- and multi-modal generative AI models that operate over structured and unstructured data, medical imaging or multi-omics data. We introduce current generative AI methods, review their applications in each data modality and discuss their strengths and weaknesses compared to traditional methods.

Results: We follow the PRISMA guideline and review 155 articles on generative AI applications in time series healthcare data across modalities. Furthermore, we offer a systematic framework for clinicians to easily identify suitable AI methods for their data and task at hand.

Conclusions: We review and critique existing applications of generative AI to time series health data with the aim of bridging the gap between computational methods and clinical application. We also identify shortcomings of existing approaches, and highlight recent advances in generative AI that represent promising directions for healthcare modeling.

(JMIR Preprints 30/04/2024:59792)
DOI: https://doi.org/10.2196/preprints.59792

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Generative AI models in time varying biomedical data: a systematic review

Rosemary He[1,2], Varuni Sarwal[1,2], Xinru Qiu[3], Yongwen Zhuang[4], Le Zhang[5], Yue Liu[6], Jeffrey N Chiang[1,2,7], Eleazar Eskin[1,2]

1. Department of Computer Science, University of California Los Angeles, CA 90095, USA
2. Department of Computational Medicine, University of California Los Angeles, CA 90095, USA
3. Division of Biomedical Sciences, School of Medicine, University of California Riverside, Riverside, CA 92521, USA
4. Department of Biostatistics, University of Michigan, MI 48109, USA
5. Institute for Integrative Genome Biology, University of California Riverside, Riverside, CA 92521, USA
6. Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, TX, 78712, USA
7. Department of Neurosurgery, University of California Los Angeles, CA 90095, USA

Corresponding Author:
Jeffrey N Chiang, PhD
Departments of Neurosurgery and Computational Medicine
David Geffen School of Medicine at UCLA
300 Stein Plaza, Suite 560
Los Angeles, CA 90095
United States
Email: njchiang@g.ucla.edu

Abstract

Background: Trajectory modeling is a longstanding challenge in the application of computational methods to healthcare. However, traditional statistical and machine learning methods do not achieve satisfactory results as they often fail to capture the complex underlying distributions of multi-modal health data, and long-term dependencies throughout patients' medical histories. Recent advances in generative AI have provided powerful tools to represent complex distributions and patterns with minimal underlying assumptions. These have had a major impact in fields such as finance and environmental sciences, and recently researchers have turned to these methods for disease modeling.

Objective: While AI methods have proven powerful, their application in clinical practice remains limited due to their black-box like nature. The proliferation of AI algorithms poses a significant challenge for non-developers to track and incorporate these advances into clinical research and application. In this work, we survey peer-reviewed, generative AI model papers with specific applications in time series health data.

Methods: Our search includes single- and multi-modal generative AI models that operate over structured and unstructured data, medical imaging or multi-omics data. We introduce current generative AI methods, review their applications in each data modality and discuss their strengths and weaknesses compared to traditional methods.

Results: We follow the PRISMA guideline and review 155 articles on generative AI applications in time series healthcare data across modalities. Furthermore, we offer a systematic framework for clinicians to easily identify suitable AI methods for their data and task at hand.

Conclusions: We review and critique existing applications of generative AI to time series health data with the aim of bridging the gap between computational methods and clinical application. We also identify shortcomings of existing approaches, and highlight recent advances in generative AI that represent promising directions for healthcare modeling.

Keywords: Generative AI, electronic health records, time-series forecasting

Introduction

Generative artificial intelligence (GenAI) is a family of AI models that are capable of generating synthetic samples for different modalities across text, images, audio, and video. Well-known genAI tools such as ChatGPT[1] and DALL-E[1] have revolutionized the way people view AI. Given some input, these tools are capable of generating new content that has not been seen before[2]. One potentially impactful application of genAI tools is in healthcare, as more digital health records are generated every day. Time series health data has been of particular interest to researchers for decades in tasks including tracking patient trajectories[3], treatment estimations[4], etc. Compared to cross-sectional data which deal with one timepoint, time series provide more information on patient trajectory and development over time. There are two main types of time series data: real time and longitudinal. While real time data usually spans a shorter period of time (minutes or days) and has more time point entries, longitudinal data refers to data recorded over the magnitude of months or years with fewer samples. As the amount of data available continues to grow, genAI models have begun to outperform traditional statistical and machine learning (ML) methods in various tasks[5]. These tools offer new ways to manage and understand patient data patterns over time and can assist clinicians in making treatment decisions across different data modalities. In

(Figure 1), we illustrates different modalities that GenAI has been applied to, including structured health records (EHR: e.g., demographics, laboratory tests, vital signs, medications, and diagnoses), images from medical scans[6, 7], and free-text, including progress and discharge notes, as well as reports from radiology and surgeries. GenAI also has applications in multi-omics[8], which looks at different molecular data to get a clearer picture of biological processes. However, introducing GenAI into healthcare comes with its set of issues. Concerns about keeping patient data private, making the AI's decision-making process transparent, needing large amounts of resources and ethical issues are all significant challenges that must be tackled. Nonetheless, GenAI tools have the potential to improve disease prediction, tailor treatments to individual patients, and make healthcare more efficient.

Relative work: existing reviews and surveys

To get a list of existing surveys on the topic of generative AI for disease trajectory, we conducted a preliminary search of review papers written after 2010 in databases including Google Scholar, Scopus and PubMed. We identified 39 entries broadly covering the use of AI in healthcare (Supp. Table 1). From a review of these works, although we see an increasing number of AI-related review papers in the biomedical setting in recent years (Supp. Figure 1), there remains a gap in reviewing generative AI approaches working with time series data specifically, prompting the need for our paper.

Algorithms and techniques overview

In this section, we introduce a set of traditional, machine learning (ML), deep learning (DL) and generative AI models. Despite a shared nomenclature, it is important to differentiate between modern AI and the field of artificial intelligence. Modern AI typically refers to advances in deep learning, which are a specific class of machine learning algorithms. Machine learning is itself a subset of the general study of artificial intelligence. Previous ML approaches have been mostly *discriminative*, such that given a set of input predictors, they learn a strategy (decision boundary) to fit or separate data points. On the other hand, GenAI, which are *generative* ML methods, capture the underlying data generation process. By learning the distribution domain of the data, these models can synthesize hypothetical data points that are statistically indistinguishable from the originals. We briefly describe methods that are referenced later and show a timeline for when these models were first introduced in (Figure 2), and a comparison among these methods using common metrics in (Table 1).

Benchmark models

First we introduce some baseline statistical models commonly used as benchmarks in modeling time series data. The Naive2 model predicts the outcome value at a certain time point by either using the observed value from the most recent observation or from a similar period in the past. For autoregressive methods, the Auto-Regressive Moving Average (ARMA) model[9] predicts the outcome value at a given time point using both a linear combination of multiple past observations and a linear combination of past error terms in the regression model. As an extension from ARMA, the Auto-Regressive Integrated Moving Average (ARIMA) model[10] accounts for seasonal or periodical changes in the data. Lastly, the exponential smoothing (ETS) model[11] uses a weighted combination of multiple past observations where the decay of the weights is exponentially distributed.

Deep learning models

Deep learning models are the combination of an architecture, i.e. the structure, and an objective, i.e. the loss function. Whether DL models are discriminative or generative is governed by the choice of objective function. Therefore, they are not inherently generative nor discriminative, and are the building blocks of modern AI models, though they have typically been developed and demonstrated in the discriminative setting. Here we introduce the canonical DL architectures which have been applied across different GenAI settings, and in the next section we introduce genAI specific models.

Convolutional Neural Networks

Convolutional Neural Networks (CNNs)[12] are a type of deep learning neural network architecture that is particularly well-suited for processing and analyzing image data. CNNs are good at recognizing patterns in small areas and work well for short-term connections due to their ability to perform convolutions step-by-step over time[13]. CNNs can be used for univariate time series forecasting, in which the trend can be viewed as a 2D graph. In addition, researchers have explored different CNN architectures for time series forecasting, such as temporal convolutional networks[14] that are tailored for sequence-to-sequence modeling tasks, leveraging the strengths of CNNs while addressing the challenges of working with sequential data.

Recurrent Neural Networks

Recurrent Neural Networks (RNNs)[15] are one of the earliest deep learning frameworks that are designed to capture sequential dependencies and patterns across time steps. RNNs have an internal/hidden state that captures information from previous time steps and influences the processing of subsequent inputs. During training, sequential data is fed into the network one step at a time. The network processes each input along with its corresponding hidden state, updating the hidden state based on the current input and the previous hidden state. This process is repeated for each time step in the sequence, allowing the network to capture temporal dependencies and learn patterns in the data. The architecture of RNN consists of multiple recurrent layers, each containing recurrent units. There are two main types of units: Long Short-Term Memory (LSTM) cells and Gated Recurrent Units (GRU). These units are responsible for capturing temporal dependencies and encoding information from previous time steps. By training on sequential data and leveraging their recurrent structure, RNNs performed strongly against traditional deep learning networks in modeling complex patterns in time series data. However, RNNs can be time consuming as inputs are processed sequentially and do not model long range dependencies well as the information diminishes over time.

Transformer

Transformers[16] are the foundation of many state-of-the-art generative models including ChatGPT. Unlike RNNs, which process data sequentially and can be computationally expensive, transformers rely on a mechanism called self-attention[16] to capture dependencies simultaneously and over long ranges. They consist of an encoder-decoder architecture, where the encoder processes the input sequence, and the decoder generates the output sequence. During training, the input is embedded into a high-dimensional space and the encoder applies a self-attention mechanism to capture relationships between different parts of the input sequence in the embedding space. Self-attention allows each position in the sequence to attend to all other positions, enabling the model to learn contextual representations efficiently. These representations are then passed through feedforward neural networks within each layer to further process the information. The decoder, on the other hand, predicts the output sequence step by step based on the encoder's contextualized representations and previous outputs. Transformer models typically include multiple layers of encoder and decoder blocks, each containing self-attention mechanisms and feedforward neural networks. In the age of

large data, transformers are one of the most powerful models to capture complex relationships in sequential data.

Generative AI models

Generative Adversarial Networks

Generative Adversarial Network (GAN)[17], is a generative vision model that consists of two parts: the generator that creates fake data, and the discriminator that critiques them. The generator's goal is to produce realistic outputs from random noise that resemble realistic images. On the other hand, the discriminator's role is to differentiate between real data and the data generated by the generator. Through this back-and-forth process of creation and critique, GANs improve over time, producing increasingly realistic results. GANs are one of the earliest image generation models that have shown great success in generating synthetic images, and have been used extensively for generation of new dataset and domain transfer in medical imaging. However, training GANs can be challenging due to issues like limited type of data produced and instability during training.

Variational Autoencoders

Autoencoders are a class of neural network methods that learn a low-dimensional representation of high-dimensional structured data. Autoencoders consist of two parts: an encoder that projects high dimensional data into a latent space with lower dimensions, and a decoder that learns to map a point in the latent space back to its high dimensional representation. However, the latent distribution of a vanilla autoencoder is unknown, making inference difficult and prompting the need for variational autoencoders (VAE). VAE models[18] make further assumptions about the sample generation process that allow for us to model an approximation of the distribution of the latent space, usually a Gaussian with a mean and variance estimated by the model. Unlike other image generation models, VAEs provide a latent space that can be estimated efficiently and used for further modeling. As with many medical applications, conditional variables offer additional information and can improve parameter estimation. A natural extension to the VAE is Conditional VAE, which includes additional conditional variables.

Diffusion Models

Denoising Probabilistic Diffusion Models (DPDM)[19] are a class of vision models that generate a new image by removing noise gradually from a pure noise input. The model learns to denoise data by understanding the probabilistic relationships between noisy and clean data points, making it particularly effective in scenarios where data is corrupted by noise. During training, DPDM simulates a diffusion process, where noise is gradually added to the data, and then iteratively removing this noise to reconstruct the original signal. At each step, the noisy observation is generated by adding a small amount of noise to the previous observation. The diffusion process is typically modeled using an autoregressive process, where the next observation is conditioned on the previous observation and noise.

Natural Language Processing Models

Natural language processing (NLP) models are a class of generative AI models that deal with unstrcutred text data, such as spoken words and medical notes. In the general domain, transformer-based[16] NLP models have achieved state-of-the-art performance for name entity recognition, relation extraction, sentence similarity, natural language inference, and question answering. Currently, Bidirectional Encoder Representations from Transformers (BERT)[20], has become state-of-the-art, achieving the best performance across many NLP benchmarks[21, 22]. Typically, these models are trained in two stages: language model pretraining (i.e., learning using a self-supervised

training objective on a large corpus of unlabeled text) and fine-tuning (i.e., applying the learned language models solving specific tasks with labeled training data).

Foundation Models for medicine

Foundation models (FMs)[23] are machine learning models capable of performing various generative tasks after being trained on extremely large, and typically unlabeled datasets[24]. In the past few years, FMs have received significant attention given their impressive range of capabilities across multiple domains. FMs trained on electronic health records have shown the ability to predict the risk of 30-day readmission[25], select future treatments[26], and diagnose rare diseases[27]. There are two broad categories of foundation models built from electronic medical record data: Clinical Language Models (CLaMs) and Foundation models for EMRs (FEMRs). The first category of FMs are Clinical Language Models, or CLaMs, which are a subtype of large language models (LLMs). The unique attribute that separates CLaMs from general LLMs is their specialization on clinical/biomedical text– CLaMs are primarily trained on, and output clinical/biomedical text. The second class of clinical FMs are Foundation models for Electronic Medical Records (FEMRs). These models are trained on the entire timeline of events in a patient's medical history. Given a patient's EMR as input, a FEMR will output not clinical text but rather a machine-understandable representation for that patient. This representation, also referred to as a patient embedding, is typically a fixed-length, high-dimensional vector which condenses large amounts of patient information. Almost all CLaMs trained on clinical text used a single database: MIMIC-III, which contains approximately 2 million notes written between 2001-2012 in the ICU of the Beth Israel Deaconess Medical Center[28]. Unfortunately, the exceptions are the very CLaMs that seem to have the best performance -- ehrBERT[29], UCSF-Bert[30], and GatorTron[31]-- as they were trained on private EMR datasets. Researchers developed BioBERT[32] (with 110 million parameters) and PubMedBERT[33] (110 million parameters) transformer models using biomedical literature from PubMed. NVIDIA developed BioMegatron models in the biomedical domain with different sizes from 345 million to 1.2 billion parameters using a more expansive set of PubMed-derived free text[34]. ClinicalBERT[35] was one of the initial largest transformer models using clinical narratives. ClinicalBERT has 110 million parameters and was trained using 0.5 billion words from the publicly available Medical Information Mart for Intensive Care III (MIMIC-III, cite) dataset. The largest clinical model so far is GatorTron, which was trained using a corpus of >90 billion words from UF Health (>82 billion), Pubmed (6 billion), Wikipedia (2.5 billion), and MIMIC III (0.5 billion).

Table 1. Metric comparison between methods mentioned (all categories out of 5 where 1 is the lowest and 5 is the highest)

Model	Computational Cost	Interpretability	Model size	Data requirement	Accuracy
Benchmark models	1	5	1	1	2
CNN	2	3	2	3	3
RNN	2	3	2	3	3
Transformer	3	4	3	3	5
GAN	4	2	4	4	5

VAE	2	3	3	4	4
Diffusion	2	1	3	4	5
NLP	3	2	3	3	4
Foundation Models	5	1	5	5	3

Methods

In this review, we survey GenAI applications in time series healthcare data and its ability to improve patient care and diagnosis process, while considering the challenges and showcasing real-world uses that highlight its importance in the field. For the selection process, we follow the PRISMA guideline and searched in PubMed and Google Scholar for conference and workshop papers, with keywords including generative AI, healthcare, time series, longitudinal, EHR/EMR and genetics. The initial search returned over 3500 papers, which we then apply the exclusion criteria as follows: peer reviewed, published after 2000, generative AI method, time series or longitudinal data for structured and imaging data, ultimately yielding 155 studies as of April 2024. We organize the review by data modality and offer suggestions on how to choose the appropriate models for each section, for clinicians who are interested in incorporating GenAI into their practice. In (Figure 3), we organize the workflow for choosing an appropriate model for the data and computational resources at hand.

Results

Structured text data

In an effort to improve patient safety and reduce medical costs, as of 2015, over 500 000 US physicians and almost 6000 hospitals should have operating electronic health records and health information technology systems[36], which a digital version of paper charts[37]. As a result, hospitals and medical practices in the United States have rapidly adopted EHR systems, resulting in massive growth of electronic patient data. Among these, the most abundance data is text, which can be structured and unstructured. We first discuss structured data, also referred to as tabular data, which typically include numeric (e.g., laboratory measurements, vital signs) and categorical (demographics, medications, diagnostic codes) features. In disease trajectory prediction, vital and lab data are the most common time series data. These are typically represented as a 2D matrix with each row as an encounter or patient, and each column as a feature of interest. Time series inputs can be represented as multiple features, such as BP_time1, BP_time2, etc, or as stacked 2D matrices where each matrix is a time point. Depending on the recorded values, feature values can either be numerical or categorical.

Applications

Generative AI has been used extensively in tabular data and we will introduce existing work in three main categories: data augmentation and imputation, disease classification and prediction, and counterfactual estimation.

EHR datasets often face the problem of missing or unavailable data, which is especially common in lab data, where not every patient in the cohort will receive the same tests. In addition, incomplete documentation or patient privacy concerns could also limit access to the complete dataset. However, naive approaches such as filling missing data with 0 or forward filling in time add noise to the dataset at the cost of performance, prompting the need for better methods. Generative

models learn to impute missing values by learning patterns from existing data and generating plausible values for missing entries, thereby enhancing the completeness and usability of the dataset. Another challenge is scarcity, where the prevalence for certain rare diseases are so low, it is impossible for models to learn the pattern. Generative models have been used to generate synthetic patient records that mimic real-world data distributions. This can help in augmenting limited datasets and improving the robustness of predictive models trained on EHR data. To impute data, a general approach will start with a RNN model with LSTM layers[38-43], or GRU layers[44]. In more advanced approaches, researchers can modify the architecture parameters based on the dataset size or augment the model with additional pieces. In this work, the authors combine the GRU with ordinary differential equations to model irregularly sampled data and impute the entire time series[45]. ImputeRNN[38] is a method that imputes data with consideration of medical bias by applying a mask during training. One can also combine deep learning methods, where in this work a variational autoencoder is used first to find a data representation, which is then passed to a RNN for imputation[46]. For filling missing data, it may be favorable to consider both previous and later observations. BRITS[47] imputes the missing data bi-directionaly with an RNN graph. While RNN is the most common method in the space, GANs have also become a popular choice. In these works, time series data is treated as an “image”, on which a GAN is trained to fill in the missing “gaps”[48-51].

The most common goal in modeling disease trajectory is for disease prediction and classification. Generative models can learn temporal dependencies within these sequences and generate realistic patient trajectories over time, enabling predictive modeling of disease progression or treatment outcomes. There are several types of predictions that can be made, a binary classification of whether an event will occur (mortality, sepsis, etc), a probability prediction of an event occurring at some time, or a time that an event occurs (survival analysis). Due to time series data’s nature of unfixed length and frequency, many ways are proposed to handle such data. In earlier approaches, time points in time series data are treated as separate entries with no relation to each other and passed through a fully connected neural network to predict the disease of interest. However, this approach ignores temporal information that can be crucial in future predictions. One of the most common ways of handling input with variable length is to first convert the input to some fixed length representation, which is then passed through a network. In this paper, measures such as body mass index, smoking status and cholesterol are first converted to an embedding of fixed size, and then passed through an RNN with LSTM layers to predict cardiac dysrhythmias[52]. In another approach, time series data including labs and vitals are divided into 5 relative timepoints to the duration of the hospital stay to fix the input length and predict mortality in heart failure patients[53]. Once the embedding choice is made, the subsequent model of choice is similar between many works. In these work, RNN models are used to predict MRSA positivity[54], mortality[44, 55-57], readmission[58-61], next diagnosis[62-65], length of stay[66, 67], hypertension[68, 69], treatment response[70], and survival analysis[71]. Further, these general use prediction models[26, 72-74] all use a RNN model to predict clinical events based on time series data. Extending from a single prediction point, this work uses LSTM to model irregularly timed time series data and predict multi-step future trajectories on diabetes and mental health cohorts[74]. In recent years, researchers have also turned to transformer models for prediction tasks[75-77] as they capture more long-range dependencies and provide more interpretability of model weights than RNNs. In this work, self-attention is used to measure each feature’s relevance to each other and also each time point’s relevance within each feature to predict delirium in critical care[78], and another paper uses similar mechanisms to predict cardiac patient mortality risk[79]. Transformers have also been useful in making multi-point predictions. BEHRT is one of the earlier models using transformers to predict clinical events, and can predict up to 300 events simultaneously[80]. DuETT is a general purpose transformer based method that takes in time series features[81]. Besides RNN and transformers, a deep diffusion model can also be used to learn complex representations of disease networks that can

be used for downstream prediction tasks[82].

Lastly, counterfactual estimations aim to track the “what-if” trajectory where if a patient was not given any medication or treatment, what would their trajectories look like. To estimate a trajectory that is never observed, one must take into account both observed and hidden confounding variables, which is challenging in causal inference models. AI models offer the advantage of making no prior assumptions about the relationship between variables, only to learn purely from data. However, it does not tackle the hidden variables problem, which is still an active problem in the field. Most commonly, RNN is used to estimate treatment outcomes and therefore provide counterfactual estimation in these works[70, 82-84]. Depending on the task and data at hand, modifications can be made to the basic RNN to achieve better results. In this work, a 2-layer bidirectional LSTM model is used for survival estimation[85]. Similarly, this work combined RNN with g-computation, a causal inference method to estimate treatment effects[86]. A combination of reinforcement learning and RNN can also be used to predict counterfactual evaluation, in this example for public intervention to COVID-19[87]. Besides RNN models, autoencoders have been used to discriminate patient characteristics from treatment to estimate counterfactuals[88].

Comparison with standard methods

Although direct comparison bewteen GenAI methods and traditional statistical methods in healthcare settings is limited, studies have been conducted to compare the performance of deep learning methods to the standard statistical methods using more general time series data.

Cerqueira et al compared machine learning methods in time series forecasting (Rule-based, random-forest, Gaussian Process regression) to statistical methods (ARIMA, naive2, ETS) using datasets with varying sample sizes up to 1000 and showed that machine learning methods out-performed statistical methods when the sample size was relatively large[89].

Makridakis et al jointly evaluated the accuracy of deep learning methods and popular statistical methods using 1045 M3 dataset[90]. The statistical methods evaluated included the Naitve2 method, ARIMA, ETS and an ensemble approach of ARIMA and ETS. The DL methods included DeepAR, FeedForward, Transformer, WaveNet and an DL ensemble method. The authors showed that among the non-ensemble approaches, the statistical methods on average yielded better performance than DL methods for short range prediction with regard to symmetric mean absolute percentage error (sMAPE) and mean absolute scaled error (MASE), while the DL methods performed better in general for medium and long range predictions[85]. The DL ensemble method using a combination of the four DL methods consistently out-performed the non-ensemble methods in the study[85].

Model selection

Unstructured text data is the most common data form in EHR and has the most existing work among all data modalities and the most number of models available, as illustrated in the workflow in (Figure 3). The flowchart is organized as sections in this paper, where different colors refer to different data modalities for easier visualization. Besides the statistical methods mentioned above, there are also a number of ML models including logistic regression, decision tree, random forest and XGboost methods (colored red in Figure 3). While we do not discuss these methods in this review, there exists review papers that describe them in detail[91-93]. If the dataset size is quite small, a traditional autoregressive or a simple ML method would work better as more complex models require more data. If the dataset size is large, however, deep learning models would perform better as the relationship between features are more complex and tend to be nonlinear. Commonly, the model of choice will be RNN or transformer-based, which also have the most literatures available. Once one has chosen a base model, they can start with a published framework and modify certain parameters based on their goal. If the model is quite simple and only consists of linear layers, they may also consider changing part of the structure to newer architectures such as attention layers, or chaining

two models together to improve performance.

Unstructured text data

In this section, we discuss unstructured data, including clinical narratives such as progress notes, radiology reports, patient correspondence and discharge summaries. Unstructured data comprise 80% of EHR data[94]. The loosely structured nature of typed text (also known as 'free text') is effective in day-to-day clinical workflows but presents a major challenge for automating EHR-based registries. The unstructured data may contain key patient information missing in structured data, extra information complementing structured data, or even data that may contradict information represented by structured data. The complexities of unstructured data, along with the fact that existing text mining tools and natural language processing applications have limited accuracy in extracting information from free text, have prompted some registries to ask for a manual chart review of individual patients before final inclusion in the registry. Unstructured data limits the application of automated computational phenotyping methods and increases the likelihood of low data quality (e.g., missing data) when data are extracted from structured EHR data only[95, 96].

Applications

Since the applications of LLMs on healthcare time series data is still a developing area of research, here we discuss static time applications of these models. We will discuss potential research directions for time series unstructured text based on progress in other fields such as climate and finance in the future works section. Both traditional NLP and foundation models can be used to perform clinical NLP tasks such as clinical concept extraction (or named entity recognition NER), medical relation extraction (MRE), semantic textual similarity (STS), natural language inference (NLI), medical question answering (MQA) and Medical Report Summarization. The domain of medical report summarization is a popular area of research with numerous papers published on the use of NLP and foundation models to summarize medical reports, healthcare records and medical dialogues[97]. A radiology report is a medical document that contains the details of an imaging study (such as X-ray, MRI, etc). A Radiology report consists of three components: (1) Background section which contains the medical history of the patient, (2) Findings section which discusses the crucial observation and findings of the radiology study, and (3) Impression section which is a short summary of Findings section. The impression section is usually written by medical professionals which is a time-consuming process. The only aim of radiology report summarization is to automate the generation of this impression section. Several attention based and FM based models have been developed to improve the quality of summarization[98].

Medical dialogue summarization corresponds to the automatic generation of coherent summaries that capture medically relevant context from dialogues between patients and healthcare providers. Medical conversation summarization can help medical providers to keep a record of patient encounters and also provide the necessary context of a patient's medical history during patient hand-offs between providers. Existing studies have leveraged techniques from computational linguistics[99], NLP (PEGASUS[100]), pretrained language models and low-shot learning to collect labelled data and perform medical dialogue summarization.

Medical question answering refers to developing a technique that automatically analyzes thousands of articles to generate a short text, ideally, in less than a few seconds, to answer questions posed by physicians [101]. Such a technique provides a practical alternative that allows physicians to efficiently seek information at point of patient care. Several QnA systems have been developed using NLP[102], Foundation Models (GMAI, MedLM).

Comparison with standard methods

The use of foundation models has several advantages and disadvantages compared to traditional

machine learning models. Foundation models haver a significantly higher computational cost compared to other traditional ML models, requiring massive datasets and GPU requirements in training. As a result, these models are composed of millions to billions of parameters, but can achieve high predictive accuracy. However, this results in poor model interpretability and a higher latency/runtime. Here, we detail six comparison parameters are listed below:

1. Cost: foundation models are several times more expensive to train and fine tune, as compared to traditional ML models. While these costs can be lowered over multiple downstream applications, their value may take longer to realize than a smaller model developed for a single high-value task[24].
2. Privacy: data privacy and security are significant concerns with foundation models, as they may leak protected health information through model weights or prompt injection attacks[103].
3. Interpretability: interpretability is of utmost importance in the healthcare domain, and foundation models have lower interpretability as compared to traditional ML models because of a significantly larger number of parameters and more complex architectures.
4. Performance: foundation models have been proven to have an improved predictive performance on classification, regression and generative tasks based on auroc, auprc, F1-Score, and Accuracy[104].
5. Multimodality: unlike traditional ML models where the majority of evaluation tasks span one data modality, foundation models are trained and evaluated on multimodal datasets consisting of both text and images.
6. Labeled Data: once trained, foundation models require smaller labeled datasets for downstream evaluation tasks, and the zero-shot and few-shot performance is high. On the other hand, traditional ML models require several thousands of training examples for the performance to scale well.

Model selection

Foundation models hold great potential to assist clinicians for a wide range of healthcare problems. However, clinicians should be aware of the risks associated with the use of these models, and potential data leaks. They should also select which models to use based on the computational resources available to them as shown in (Figure 3), since the training of these models is getting increasingly expensive (Table 1).

Medical imaging

Medical imaging refers to 2D or 3D images obtained from radiology procedures such as magnetic resonance imaging(MRI), computed tomography(CT) scan and waveform recordings (EEG, ECG, etc.). Generative AI has been applied to a wide set of problems in medical imaging to aim and solve the problems of scarcity, heterogeneity between datasets, low resolution quality, etc. The medical images are usually first aligned to an existing template for uniformity, and are represented by 2D matrices or 3D vectors of the pixels/voxels. There are three main classes of generative vision methods: GANs, VAEs, and DDPMs. Each of these classes have been applied in medical imaging including waveform data. Among all applications, medical image series and waveform series data input have been used mostly in anomaly detection and classification, data imputation, and image segmentation and registration tasks.

Applications

The most common application in longitudinal images is anomaly detection and disease progression prediction. For clinical diagnosis, generative AI based anomaly detection tools have been developed where the model identifies deviations in one image from the rest of the population, aiding in the

detection of diseases or abnormalities. For medical image series, previous scans are used as inputs or conditional variables to predict the next image. Many works in this space are a variation of GAN, as they were introduced earlier than VAEs and DDPMs. GP-GAN uses stacked GANs to predict tumor growth from longitudinal MRI[105]. MPGAN makes longitudinal Prediction of Infant MRI With Multi-Contrast Perceptual Adversarial Learning[106]. In LD-GAN[107], the authors predict disease progression images with missing MRI in the input. In these papers, the authors train a GAN to predict autistic spectrum disorder given longitudinal sMRI scans[108]. More recently, VAEs have been applied extensively as they offer a latent space to work with. In general, researchers start with a vanilla VAE and apply slight modifications to fit their data modality and goal to detect and predict diseases including: MS[109], AD[110-113], tau-biomarker detection for AD[114], glaucoma[115] and lung cancer[116]. There has also been an increase in works published using multi-modality data such as imaging data and tabular or genetics data. In this work, the authors train a VAE model to predict disease progression score based on longitudinal images and microRNA data input[117]. In dealing with multi-modality inputs, a combination of deep learning models dealing with each modality may also be advantageous. This work combined VAE with a linear mixed-effect model to learn a Riemman progression model that can be applied to general disease trajectory estimation[118]. Another work combined an autoencoder framework with attention units in a transformer to predict final ischemic stroke lesions from MRI[119]. In this paper, a spatio-temporal convolutional LSTM is learned to combine imaging and non-imaging data to predict tumor growth images[120]. C-SliceGen is a method that takes an arbitrary axial slice in the abdominal region as a condition and estimates a vertebral level slice[121].

Similar to tabular data, data generation is also a task of interest. For imputing waveform data such as EEG and EGG signals, GANs have been used extensively[122-127], along with diffusion models[128]. For longitudinal medical image series, all three classes have been applied. In this work, a recurrent conditional GAN is used to generate sequential image data conditioned on past variables. Diffusion based models have also been applied to impute longitudinal CT and MR medical images[129]. Finally, an example of applying autoencoders with contractions on the latent space is used to predict and generate aging brain images with only the input of age and disease state[130].

A very common task in medical imaging is image segmentation and registration. We note here that while image segmentation and registration tasks are different concepts, the underlying models applied to solve the problems are largely the same. In this work, an encoder-decoder structure model is used for contrastive learning for image segmentation[131]. Extending from autoencoders, a conditional VAE is used to produce cardiac image segmentation conditioned on pre-defined anatomical criteria[132]. We note here that while work in this area with time series input is scarce, the underlying model and training process remain the same for single time input models[133-138]. Furthermore, many of the methods mentioned above for anomaly detection can also be applied to registration and segmentation tasks during training.

Comparison with standard methods

As an alternative to CNN and vision models, a common way to analyze imaging data is to convert the image to numeric outputs (by extracting measures such as thickness, mean intensity, etc) that can be fed into linear models for tasks such as classification and segmentation. The first step is considered feature extraction, which is separate from subsequent tasks. To work with images directly, linear models of estimation such as PCA or single value decomposition are often used. These methods estimate a population average and individual variabilities based on external variables and generate a new image by randomly sampling variations to be added onto the population average. However, these generations often have poor resolution or fail to adequately capture variability. In contrast, modern neural network approaches (e.g., CNN) *jointly* learn the optimal feature extraction strategy together with the downstream task. GenAI algorithms are thus able to consider highly nonlinear and nuanced interactions between inputs that better capture the underlying distribution of data.

Model selection

With computer vision models' success in recent years, there is a lot of potential in their applications in medical imaging. In recent years, diffusion models have shown great promise in generating images that are comparable to the state-of-the-art GAN methods. Furthermore, GenAI methods that support video generation can potentially be adapted for medical imaging[139-143]. However, some issues that are present in current vision models are also propagated to medical imaging, such as illogical generations that do not comply with physical laws. Therefore, clinicians should take the generations with a grain of salt when applying it to downstream tasks.

Although there are many tasks associated with imaging data input, the underlying model architecture remains largely the same for prediction, synthesis, registration or even domain transfer tasks. For anyone with time series or longitudinal images for any tasks, choosing one of the three families of vision methods following the workflow in (Figure 3) is a good place to start. One may also refer to (Table 1) to compare the models' strengths and weaknesses. GANs are an earlier framework that remain powerful today, especially with synthetic generation, but are more complicated to implement as it has a generator and a discriminator. On the other hand, VAEs offer a latent space that can be used to downstream linear models such as mixed effect models to control for variables such as age, but its generation quality may not be as ideal as the other two. Lastly, diffusion models are largely applied in recent years due to their simple architecture compared to the other two and low memory requirements, but offer very little interpretability on the generation process. For clinicians with limited computing resources, large 3D images may be too large to fit onto GPU, and in many cases the images can be split into smaller patches to reduce memory footprint[144]. Lastly, there are publicly available general purpose vision models that can be downloaded as a starting point, which can be further trained and fine-tuned for the specific task at hand.

Genetics and multi-omics data

Trajectory modeling in multi-omics data helps us understand how biological systems like cell development, diseases, or reactions to treatments change over time. Multi-omics data emcompasses genomics, transcriptomics, proteomics, epigenomics, and metabolomics, offering a detailed picture of biological activities at various molecular layers. In genomics, DNA sequences are represented as strings of nucleotides ATGC, which can be transformed into a numerical matrix where each nucleotide is represented as a binary vector. For transcriptomics, expression data is often represented as a matrix of continuous numeric values where each row corresponds to a sample or a cell and each

column to a gene or transcript; proteomics data is similarly structured but represents peptide abundances or protein expression levels, and epigenetic data such as methylation status is typically captured as binary values or continuous levels indicating methylation at specific sites. By merging this data and using trajectory modeling, clinicians and scientists can uncover time-related trends and critical factors that influence those biological shifts.

ML and DL applications in multi-omics

Identifying pathogenic mutations is pivotal for unraveling the genetic underpinnings of diseases. The task of variant effect prediction (VEP), which seeks to determine the phenotypic implications of genetic variations, stands as a cornerstone challenge in human genetics. Traditional approaches, leveraging statistical methods alongside evolutionary conservation scores such as SIFT and PhyloP, have made strides in classifying variants of uncertain significance[145, 146]. Despite these advancements, accurately differentiating between variants that disrupt protein function and those with neutral effects remains a daunting challenge. The persistence of many missense variants within the gray area of VUS significantly hampers the diagnostic utility of exome sequencing in clinical settings.

To effectively address the challenge of categorizing mutations, particularly those classified as variants of uncertain significance (VUS), ML techniques have been increasingly applied. These methods leverage a broad spectrum of features, including documented phenotypic outcomes of known genetic variants, functional annotations from genomic databases, evolutionary conservation scores that highlight the significance of sequences across species, and epigenetic data offering insights into gene regulation mechanisms. Among the array of ML methodologies, random forests, gradient boosting models, and support vector machines (SVM) stand out as potent tools for training complex genetic datasets to decipher VUS. Random forest, for example, operates by constructing multiple decision trees during training and outputting the mode of the classes (for classification) or the mean prediction (for regression) of the individual trees. This approach is particularly adept at managing the heterogeneity and high dimensionality of genetic data, with DEOGEN2, RENOVO, and VEST4 serving as prime examples of random forest application in this field[147-149]. Gradient boosting models enhance decision-making by building an ensemble of decision trees sequentially, with each new tree correcting errors made by the previous ones. This method excels at detecting intricate patterns within data, thereby producing highly precise classification models. M-CAP and MutPred2 utilize gradient-boosting trees to demonstrate the method's effectiveness in markedly reducing the number of VUS in clinical exomes, all the while maintaining high sensitivity[150, 151]. Moreover, the integration of multiple ML methods, as seen in ClinPred, which combines random forest and gradient boosting models, exemplifies the pursuit of superior performance in predicting the impact of VUS. Despite these advancements and superior performance, ML methods face limitations, particularly in handling the vast complexity and the subtle nuances inherent in genetic data. The high dimensionality of genomic information and the intricate interactions between genetic variants often exceed the capacity of traditional ML techniques, leading to potential overfitting and a lack of generalizability to unseen data.

DL methods, including CNNs and RNNs, offer a promising solution for addressing the challenges of modeling the complex, non-linear relationships found in high-dimensional biological data. These methods significantly diminish the need for manual feature engineering by efficiently learning from raw data. Among the DL models developed for predicting VUS, such as ESM1b, EVE, MetaRNN, MBP, PrimateAI, and DANN. ESM1b is particularly noteworthy for its innovative use of a protein language model developed by MetaAI[152-157]. ESM1b has demonstrated exceptional capability in classifying approximately 150,000 ClinVar/HGMD variants and annotating around 2 million variants as damaging in specific protein isoforms, a feat not previously achieved with ML-based methods. This focus on isoform-specific annotations underlines the critical importance of considering all protein isoforms in variant effect predictions, marking a significant advancement that

ESM1b contributes to genetic research.

Trajectories modeling in multi-omics

Time-series gene expression data has been an abundant source for studying the complicative of biological processes and disease development. Huang et al.[158] analyzed time-series expression data from infected human volunteers with influenza. They identified distinct temporal patterns of gene expression that could discriminate between asymptomatic and symptomatic infections. Similar approaches have been used to study the response to various treatments, such as interferon-beta therapy in multiple sclerosis patients[159].

Single-cell transcriptomics data has revolutionized our understanding of cellular dynamics. Pseudotime inference is a widely used approach for reconstructing cellular trajectories from single-cell data. The idea behind pseudotime is to order cells along a continuous trajectory based on their transcriptional similarity. Monocle uses single-cell variations to sequence cells in pseudotime, depicting their progression through biological processes like differentiation, based on gene expression. Monocle2[160] employs reversed graph embedding (RGE) with a minimum spanning tree algorithm for pseudotime reconstruction, while Monocle3[161] enhances this with principal graph learning for refined trajectory inference. Wanderlust adjusts each cell's trajectory position using weighted averages from the shortest-path distances of randomly chosen waypoints until convergence, thereby producing an average trajectory. scVelo[162], a likelihood-based dynamical model, estimates RNA velocity to derive dynamic insights from RNA sequencing data, analyzing gene-level transcriptional dynamics by determining gene-specific transcription, splicing, and degradation rates, suitable for both transient states and systems with varying subpopulation kinetics. SCUBA (Single-Cell Universal Bayesian Analysis)[163] forecasts the temporal evolution of gene expression in single cells using theories of nonlinear dynamics and stochastic differential equations, aiding in the comprehension of gene expression dynamics. PAGA (Partition-based Graph Abstraction)[164] merges clustering and trajectory inference in single-cell RNA sequencing data, creating a graph to depict cell relationships based on gene expression and enhance understanding of cellular transitions. Lastly, TIGON[165] uses dynamic unbalanced optimal transport (OT) based on the Wasserstein-Fisher-Rao (WFR) distance to integrate temporal datasets, which provides a framework for connecting temporal measurements and predicting novel dynamics.

In the fields of proteomics, epigenomics, and metabolomics, CellTag-multi[166], a method for single-cell lineage tracing across scRNA-seq and scATAC-seq assays, mapped transcriptional and epigenomic states of progenitor cells, significantly improved cell fate prediction. Palii, Carmen G., et al.[167] used single-cell proteomics to define the temporal hierarchy of human erythropoiesis, the analysis reveals a timely ordered appearance and disappearance of transient cell populations or stages that accumulate at various positions along the erythroid trajectory, with cells undergoing gradual transitions between these stages. Additionally, Linear Mixed Models (LMMs) have been utilized to quantify temporal changes in metabolite concentrations, offering insights into the dynamics of the metabolic system[168, 169].

GenAI in multi-omics and model selection

GenAI has shown great promise in analyzing multi-omics data for disease progression analysis. By utilizing techniques such as omicsGAN, which employs Generative Adversarial Networks (GANs), and GLUE and MultiVI, which use VAEs, GenAI facilitates the integration and compression of multi-omics data from diverse sources into a unified representation[170-172]. This aids in a more comprehensive analysis of the data. Additionally, tools like PRESCIENT leverage generative modeling to map out potential landscapes from time-series single-cell transcriptomics data, making it

possible to generate trajectories for unseen data points[173]. This opens up avenues for hypothesizing about biological perturbations and pathways of disease progression. Moreover, GenAI proves instrumental in identifying distinct disease subtypes and stratifying patients based on their multi-omics profiles. For example, Gene-SGAN utilizes a GAN to generate imaging features from brain MRIs and SNPs to explore conditions such as Alzheimer's Disease and hypertension [174], while the T-GAN-D model stratifies breast cancer patients into high vs. low-risk categories based on their transcriptome profiles[175]. For clinicians interested in working with multi-omics data, we outline a procedure for finding the best tool in (Figure 3) based on their specific needs, such as resource availability and the type of data analysis required, including genetics, single-cell transcriptomics, and proteomics/epigenomics.

Discussion

Challenges and ethical considerations

Regulatory and Legal Challenges

Policies such as the Health Insurance Portability and Accountability Act (HIPAA) in the U.S.[176], along with the General Data Protection Regulation (GDPR) in the European Union[177], provide regulatory frameworks to safeguard personal health information. However, the advent of new genAI technologies in healthcare presents a challenge to these existing laws. While these regulations clearly address the handling of directly identifiable health data, the nuances of AI-generated synthetic data or the indirect identification risks posed by AI analyses can be less clearly defined[178]. The rapidly evolving nature of genAIs further complicates compliance. Traditional regulatory frameworks are often static and slow to adapt to technological advancements. In contrast, AI in healthcare is rapidly evolving, with new models and applications emerging at a pace that surpasses regulatory updates. The surge in new genAI models since 2019, is reflected by the exponentially growing number of published articles[179]. This mismatch can result in healthcare providers and AI developers being unsure about their legal responsibilities and the best practices for protecting patient privacy when using these new technologies. For instance, chatbots like ChatGPT generate responses that may contain portions of copyrighted materials, such as books or novels, and potentially include personal privacy information, such as addresses or passwords[180, 181].

Addressing these legal challenges requires a concerted effort from policymakers, legal experts, AI developers, and healthcare providers. Together, they must work to update existing laws, formulate new regulations, and establish ethical guidelines that ensure the safe, responsible, and equitable use of generative AI in healthcare[182]. So far, some progress has been made. The U.S. congressional leaders are currently formulating the SAFE Innovation Framework to steer AI legislation, with a focus on security, accountability, and explainability[183]. The world's first major act to regulate AI was passed by European lawmakers on March 13, 2024, to ensure the safety and compliance of AI[184]. In 2023, the Cyberspace Administration of China issued the Interim Measures for the Administration of Generative Artificial Intelligence Services (PRC AI Regulations), emphasizing the responsible and transparent utilization of Generative AI Services[185].

Technical Challenges

Using pre-training and fine-tuning for time series forecasting in healthcare, such as predicting patient outcomes, faces two main problems: the short length of time series and significant changes in data over time[186]. The shortness of healthcare time series makes it hard to train models because there's not much historical data on patients, increasing the risk of models learning from random noise instead of real trends and reducing their ability to make general predictions. This lack of data also makes it difficult to estimate parameters accurately, limiting how well models can understand and

predict health trends. Additionally, changes in patient conditions, treatment methods, or external factors over time can alter the data distribution, requiring models that can adapt to these new patterns to ensure they continue to forecast accurately across various conditions and patient groups[187].

Ethical Considerations

While generative AI offers significant potential in healthcare, particularly in understanding and predicting disease trajectories, its implementation must be approached with caution for ethical considerations[188], such as privacy, security, bias, and transparency.

Data Ownership

In the complex terrain of healthcare generative AI, the notion of data ownership emerges as a pivotal concern, interweaving regulatory, ethical, and practical dimensions. At its core, data ownership refers to the rights and responsibilities associated with the control, use, and dissemination of data[189]. In the healthcare sector, this encompasses a broad spectrum of data types, from patient medical records to genetic information and AI-generated synthetic datasets. The ambiguity surrounding ownership of such data poses significant challenges, particularly in light of new AI technologies that defy traditional categorizations of data generation and utilization. For instance, the generation of synthetic data by AI for research or diagnostic purposes raises questions about who holds the rights to this data —the entity that developed the AI, the healthcare providers who supplied the original data, or the patients themselves.

Data Privacy

Data privacy in the context of healthcare generative AI pertains to individuals' right to control how their personal information is collected, used, and shared[190]. Current regulations mandate the adherence to principles of data minimization, purpose limitation, and consent, among others, ensuring that patient data is handled in a manner that respects privacy rights[191].Despite existing regulations, the implementation of generative AI technologies presents intricate challenges in preserving patient confidentiality. Primarily, the character of generative AI, especially in the context of training LLMs, requires extensive patient data during the training phase[192]. In health care, training such models implies an enormous amount of detailed patient histories, genetic information, and other sensitive data points will be used. Re-identification of anonymized data via advanced algorithms presents a substantial risk to data privacy, potentially subjecting individuals to privacy violations and unauthorized access to their personal health records[193, 194]. Consequently, it is imperative that raw patient data is adequately protected and cannot be retrieved by any means in the final product.

Data Security

As AI models grow larger, the need for healthcare data and collaboration grows and with it the number of healthcare data/model breach events[195]. In 2023 alone, over 500 cases affecting more than 112 million individuals were reported[196, 197]. At the time of writing this review, 120 new cases were reported on the OCR website within the first two months of 2024. Unauthorized access or breaches to the training data could result in the disclosure of highly sensitive and confidential patient information, violating HIPPA and other confidentiality laws[198]. Phishing, malware, and data center penetration represent conventional cyberattack methodologies. However, generative AI, particularly LLMs, have introduced another form of attack known as prompt injection. This issue arises when an end user manipulates the model by inputting instructions that compel it to circumvent its pre-established operational parameters, including safety and content restrictions. Such manipulation can result in the model executing actions or producing outputs that it was programmed to avoid, such as engaging in prohibited discussions or disclosing sensitive information[199]. As we continue to

explore the potential of generative AI in healthcare, prioritizing data security will be key to achieving its full benefit while respecting the rights and dignity of patients. Currently, the most comprehensive countermeasures employed include prompt filtering and the human-in-the-loop system[199, 200]. However, these measures come at the price of poorer model performance, prompting the need for a better AI-suited strategy. In recent work, Salem et al. implemented Maatphor to mark prompt injection datasets at the pretraining phase[201], Piet et al. employed Jatmo to defend against prompt injection during the training phase[202], and Chen et al. utilized StruQ to eliminate the harmful portion of the prompt during the filtering phase[203]. As the field grows rapidly, more strategies will be developed to combat attacks on data security.

Bias

Another critical issue of generative AI in healthcare is the potential for bias and discrimination. These models are trained on vast amounts of data, which may inadvertently encode societal biases, stereotypes, or historical patterns of discrimination[204]. Such biases in AI can lead to disparities in treatment across racial, gender, or patient groups[205]. A notable example is an algorithm flagged by Obermeyer et al. for racial bias in patient care prioritization in the US, favoring white patients due to its reliance on healthcare costs as a health proxy[206]. Additionally, Zack et al. found that GPT-4 might reinforce gender biases by associating diseases with specific genders[207]. The root causes of AI bias are multi-layered, spanning from data collection to model development, leading to underperformance for marginalized groups[208-211]. These include data collection flaws, historical data inequalities, and subjective data annotation. To address these pervasive issues, a comprehensive and informed strategy is required across the entire spectrum of AI development and implementation processes.

Clinical Safety and Reliability

The assessment of clinical safety and reliability is paramount for healthcare genAIs. It underpins the trust and efficacy of these technologies in medical settings. One challenge is the potential for generative AI models to produce inconsistent, contradictory, or erroneous, particularly when dealing with complex medical scenarios or rare conditions[212]. One cause is that these models may exhibit hallucinations, especially with LLMs. They generate plausible-sounding but factually incorrect information, or make recommendations that could potentially harm patients if acted upon without proper scrutiny[213]. In healthcare, the tolerance for hallucination is considerably lower and more stringent. A single error could result in delayed or inappropriate treatment, potentially exacerbating the patient's condition or causing unnecessary harm.

Transparency and Explainability

The "black box" nature of AI models, particularly in deep learning models, makes transparency and explanability difficult to achieve. Transparency refers to the openness and accessibility of AI systems[214]. Explainability concerns the capacity of AI systems to offer comprehensible and interpretable justifications for their outputs or decisions[215]. Clear and confident decisions are critical in healthcare, and the obscurity of AI processes poses a significant obstacle to its acceptance and utilization[216]. Efforts to enhance AI's transparency and explainability are increasing. Explainable AI (XAI) techniques, such as feature importance scores, decision trees, and neural network visualization activations are often utilized[217]. Regulatory and ethical frameworks also play a crucial role in promoting transparency and explainability and need to be better established[218].

Future Directions

In this section, we highlight some fundamental advances in GenAI that could impact healthcare and discuss possible future directions of the field. Large vision models have moved beyond static imaging and have shown the ability to generate and extrapolate videos. Most notable of which, Sora from OpenAI is a model that can generate realistic videos from text inputs[219]. Such capabilities could be used to better characterize and forecast imaging-based trajectories. Similarly, foundational models for time series data are a developing area of research, with several technical papers recently published. One of the most notable, TimeGPT[220] has been developed as a general forecasting model and exceeded state-of-the-art across multiple domains, including finance, web traffic, IoT, weather, demand, and electricity. The transition from discriminative to generative modeling enabled by GenAI also pushes the capabilities of personalized medicine, for example individual treatment effect estimation and prognostication via trajectory generation[221].

Current applications of GenAI are also limited in that they are mostly restricted to the inpatient setting, where time-varying data are regularly sampled with high fidelity. By contrast, outpatient data are sparsely sampled and heterogeneous. The generative capabilities of GenAI mentioned above also provide a path for model-based inference in the outpatient setting.

Conclusion

While GenAI continues to be applied to healthcare, its application to time series healthcare data is still limited. In this survey we cover and critique existing applications of GenAI to time series health data, noting that recently developed technical capabilities remain to be exploited. To facilitate better collaboration between technical and applied disciplines, we hope this review bridges the gap and promotes collaboration to develop better methods that are suitable and robust in a practical setting.

References

1. Kumar S, Musharaf D, Musharaf S, Sagar AK, editors. A Comprehensive Review of the Latest Advancements in Large Generative AI Models. International Conference on Advanced Communication and Intelligent Systems; 2023: Springer.
2. Gozalo-Brizuela R, Garrido-Merchan EC. ChatGPT is not all you need. A State of the Art Review of large Generative AI models. arXiv preprint arXiv:2301 04655. 2023 2023.
3. Nguena Nguefack HL, Pagé MG, Katz J, Choinière M, Vanasse A, Dorais M, et al. Trajectory modelling techniques useful to epidemiological research: a comparative narrative review of approaches. Clinical epidemiology. 2020 2020:1205-22.
4. Guess HA, Jacobsen SJ, Girman CJ, Oesterling JE, Chute CG, Panser LA, et al. The role of community-based longitudinal studies in evaluating treatment effects: example: benign prostatic hyperplasia. Medical care. 1995 1995:AS26-AS35.
5. Bandi A, Adapa PVSR, Kuchi YEVPK. The power of generative ai: A review of requirements, models, input--output formats, evaluation metrics, and challenges. Future Internet. 2023 2023;15(8):260.
6. Arora A, Arora A. Generative adversarial networks and synthetic patient data: current challenges and future perspectives. Future Healthc J. 2022 Jul;9(2):190-3. PMID: 35928184. doi: 10.7861/fhj.2022-0013.
7. Koohi-Moghadam M, Bae KT. Generative AI in Medical Imaging: Applications, Challenges, and Ethics. J Med Syst. 2023 Aug 31;47(1):94. PMID: 37651022. doi: 10.1007/s10916-023-01987-4.
8. Chen C, Wang J, Pan D, Wang X, Xu Y, Yan J, et al. Applications of multi-omics analysis in human diseases. MedComm (2020). 2023 Aug;4(4):e315. PMID: 37533767. doi: 10.1002/mco2.315.
9. Whittle P. Some results in time series analysis. Scandinavian Actuarial Journal. 1952 1952;1952(1-2):48-60. doi: 10.1080/03461238.1952.10414182.
10. Box GEP, Pierce DA. Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models. Journal of the American Statistical Association. 1970 1970;65(332):1509-26. doi: 10.1080/01621459.1970.10481180.
11. Gardner ES, Jr. Exponential smoothing: The state of the art. Journal of forecasting. 1985 1985;4(1):1-28.
12. LeCun Y, Bengio Y. Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks. 1995;3361(10):1995.
13. Wibawa AP, Utama ABP, Elmunsyah H, Pujianto U, Dwiyanto FA, Hernandez L. Time-series analysis with smoothed Convolutional Neural Network. J Big Data. 2022;9(1):44. PMID: 35495076. doi: 10.1186/s40537-022-00599-y.
14. Lea C, Vidal R, Reiter A, Hager GD, editors. Temporal convolutional networks: A unified approach to action segmentation. Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14; 2016: Springer.
15. Williams RJ, Zipser D. A learning algorithm for continually running fully recurrent neural networks. Neural computation. 1989;1(2):270-80.
16. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Advances in neural information processing systems. 2017;30.
17. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. Advances in neural information processing systems. 2014;27.
18. Kingma DP, Welling M. Auto-encoding variational bayes. arXiv preprint arXiv:13126114. 2013.
19. Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. Advances in neural information processing systems. 2020;33:6840-51.
20. Devlin J, Chang M-W, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional

transformers for language understanding. arXiv preprint arXiv:181004805. 2018.
21. Wang A, Singh A, Michael J, Hill F, Levy O, Bowman SR. GLUE: A multi-task benchmark and analysis platform for natural language understanding. arXiv preprint arXiv:180407461. 2018.
22. Wang A, Pruksachatkun Y, Nangia N, Singh A, Michael J, Hill F, et al. Superglue: A stickier benchmark for general-purpose language understanding systems. Advances in neural information processing systems. 2019;32.
23. Wornow M, Xu Y, Thapa R, Patel B, Steinberg E, Fleming S, et al. The shaky foundations of clinical foundation models: a survey of large language models and foundation models for EMRs. arXiv preprint arXiv:230312961. 2023.
24. Bommasani R, Hudson DA, Adeli E, Altman R, Arora S, von Arx S, et al. On the opportunities and risks of foundation models. arXiv preprint arXiv:210807258. 2021.
25. Pang C, Jiang X, Kalluri KS, Spotnitz M, Chen R, Perotte A, et al., editors. CEHR-BERT: Incorporating temporal information from structured EHR data to improve prediction tasks. Machine Learning for Health; 2021: PMLR.
26. Choi E, Bahadori MT, Schuetz A, Stewart WF, Sun J. Doctor ai: Predicting clinical events via recurrent neural networks. Machine learning for healthcare conference; 20162016. p. 301-18.
27. Prakash P, Chilukuri S, Ranade N, Viswanathan S, editors. RareBERT: transformer architecture for rare disease patient identification using administrative claims. Proceedings of the AAAI conference on artificial intelligence; 2021.
28. Johnson AE, Pollard TJ, Shen L, Lehman L-wH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Scientific data. 2016;3(1):1-9.
29. Li F, Jin Y, Liu W, Rawat BPS, Cai P, Yu H. Fine-tuning bidirectional encoder representations from transformers (BERT)–based models on large-scale electronic health record notes: an empirical study. JMIR medical informatics. 2019;7(3):e14830.
30. Sushil M, Ludwig D, Butte AJ, Rudrapatna VA. Developing a general-purpose clinical language inference model from a large corpus of clinical notes. arXiv preprint arXiv:221006566. 2022.
31. Yang X, Chen A, PourNejatian N, Shin HC, Smith KE, Parisien C, et al. A large language model for electronic health records. NPJ digital medicine. 2022;5(1):194.
32. Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics. 2020;36(4):1234-40.
33. Gu Y, Tinn R, Cheng H, Lucas M, Usuyama N, Liu X, et al. Domain-specific language model pretraining for biomedical natural language processing. ACM Transactions on Computing for Healthcare (HEALTH). 2021;3(1):1-23.
34. Shin H-C, Zhang Y, Bakhturina E, Puri R, Patwary M, Shoeybi M, et al., editors. BioMegatron: larger biomedical domain language model. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP); 2020.
35. Huang K, Altosaar J, Ranganath R. Clinicalbert: Modeling clinical notes and predicting hospital readmission. arXiv preprint arXiv:190405342. 2019.
36. Madden JM, Lakoma MD, Rusinak D, Lu CY, Soumerai SB. Missing clinical and behavioral health data in a large electronic health record (EHR) system. Journal of the American Medical Informatics Association. 2016;23(6):1143-9.
37. Chiu C-C, Wu C-M, Chien T-N, Kao L-J, Li C, Chu C-M. Integrating structured and unstructured EHR data for predicting mortality by machine learning and latent Dirichlet allocation method. International Journal of Environmental Research and Public Health. 2023;20(5):4340.
38. Zhang J, Yin P. Multivariate time series missing data imputation using recurrent denoising autoencoder. 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 20192019. p. 760-4.
39. Suo Q, Yao L, Xun G, Sun J, Zhang A. Recurrent imputation for multivariate time series with missing values. 2019 IEEE international conference on healthcare informatics (ICHI); 20192019. p.

1-3.
40. Lipton ZC, Kale DC, Wetzel R, Others. Modeling missing data in clinical time series with rnns. Machine Learning for Healthcare. 2016 2016;56(56):253-70.
41. Shi Z, Wang S, Yue L, Pang L, Zuo X, Zuo W, et al. Deep dynamic imputation of clinical time series for mortality prediction. Information Sciences. 2021 2021;579:607-22.
42. Sangeetha M, Senthil Kumaran M. Deep learning-based data imputation on time-variant data using recurrent neural network. Soft Computing. 2020 2020;24(17):13369-80.
43. Yoon J, Zame WR, van der Schaar M. Estimating missing data in temporal data streams using multi-directional recurrent neural networks. IEEE Transactions on Biomedical Engineering. 2018 2018;66(5):1477-90.
44. Che Z, Purushotham S, Cho K, Sontag D, Liu Y. Recurrent neural networks for multivariate time series with missing values. Scientific reports. 2018 2018;8(1):6085.
45. Zaman MAU, Du D. A stochastic multivariate irregularly sampled time series imputation method for electronic health records. BioMedInformatics. 2021 2021;1(3):166-81.
46. Mulyadi AW, Jun E, Suk H-I. Uncertainty-aware variational-recurrent imputation network for clinical time series. IEEE Transactions on Cybernetics. 2021 2021;52(9):9684-94.
47. Cao W, Wang D, Li J, Zhou H, Li L, Li Y. Brits: Bidirectional recurrent imputation for time series. Advances in neural information processing systems. 2018 2018;31.
48. Gupta M, Beheshti R. Time-series imputation and prediction with bi-directional generative adversarial networks. arXiv preprint arXiv:2009 08900. 2020 2020.
49. Gupta M, Phan T-LT, Bunnell HT, Beheshti R. Concurrent imputation and prediction on EHR data using bi-directional GANs: Bi-GANs for EHR imputation and prediction. Proceedings of the 12th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics; 20212021. p. 1-9.
50. Luo Y, Cai X, Zhang Y, Xu J, Others. Multivariate time series imputation with generative adversarial networks. Advances in neural information processing systems. 2018 2018;31.
51. Li J, Cairns BJ, Li J, Zhu T. Generating synthetic mixed-type longitudinal electronic health records for artificial intelligent applications. NPJ Digital Medicine. 2023 2023;6(1):98.
52. Guo A, Smith S, Khan YM, Langabeer JR, II, Foraker RE. Application of a time-series deep learning model to predict cardiac dysrhythmias in electronic health records. PloS one. 2021 2021;16(9):e0239007.
53. Jawadi Z, He R, Srivastava PK, Fonarow GC, Khalil SO, Krishnan S, et al. Predicting in-hospital mortality among patients admitted with a diagnosis of heart failure: a machine learning approach. ESC Heart Failure. 2024 2024.
54. Nigo M, Rasmy L, Mao B, Kannadath BS, Xie Z, Zhi D. Deep learning model for personalized prediction of positive MRSA culture using time-series electronic health records. Nature Communications. 2024 2024;15(1):2036.
55. Yu R, Zheng Y, Zhang R, Jiang Y, Poon CCY. Using a multi-task recurrent neural network with attention mechanisms to predict hospital mortality of patients. IEEE journal of biomedical and health informatics. 2019 2019;24(2):486-92.
56. Ge W, Huh J-W, Park YR, Lee J-H, Kim Y-H, Turchin A. An Interpretable ICU Mortality Prediction Model Based on Logistic Regression and Recurrent Neural Networks with LSTM units. AMIA Annual Symposium Proceedings; 20182018. p. 460.
57. Caicedo-Torres W, Gutierrez J. ISeeU: Visually interpretable deep learning for mortality prediction inside the ICU. Journal of biomedical informatics. 2019 2019;98:103269.
58. Zhang J, Kowsari K, Harrison JH, Lobo JM, Barnes LE. Patient2vec: A personalized interpretable deep representation of the longitudinal electronic health record. IEEE Access. 2018 2018;6:65333-46.
59. Wang H, Cui Z, Chen Y, Avidan M, Abdallah AB, Kronzer A. Cost-sensitive deep learning for early readmission prediction at a major hospital. Canada Proc BIOKDD. 2017 2017(17).

60. Lin Y-W, Zhou Y, Faghri F, Shaw MJ, Campbell RH. Analysis and prediction of unplanned intensive care unit readmission using recurrent neural networks with long short-term memory. PloS one. 2019 2019;14(7):e0218942.
61. Barbieri S, Kemp J, Perez-Concha O, Kotwal S, Gallagher M, Ritchie A, et al. Benchmarking deep learning architectures for predicting readmission to the ICU and describing patients-at-risk. Scientific reports. 2020 2020;10(1):1111.
62. Lipton ZC, Kale DC, Elkan C, Wetzel R. Learning to diagnose with LSTM recurrent neural networks. arXiv preprint arXiv:1511 03677. 2015 2015.
63. Pham T, Tran T, Phung D, Venkatesh S. Deepcare: A deep dynamic memory model for predictive medicine. Advances in Knowledge Discovery and Data Mining: 20th Pacific-Asia Conference, PAKDD 2016, Auckland, New Zealand, April 19-22, 2016, Proceedings, Part II 20; 20162016. p. 30-41.
64. Yang Y, Zheng X, Ji C. Disease prediction model based on bilstm and attention mechanism. 2019 IEEE international conference on bioinformatics and biomedicine (BIBM); 20192019. p. 1141-8.
65. Guo W, Ge W, Cui L, Li H, Kong L. An interpretable disease onset predictive model using crossover attention mechanism from electronic health records. IEEE Access. 2019 2019;7:134236-44.
66. Zhang D, Yin C, Zeng J, Yuan X, Zhang P. Combining structured and unstructured data for predictive models: a deep learning approach. BMC medical informatics and decision making. 2020 2020;20:1-11.
67. Purushotham S, Meng C, Che Z, Liu Y. Benchmarking deep learning models on large healthcare datasets. Journal of biomedical informatics. 2018 2018;83:112-34.
68. Mohammadi R, Jain S, Agboola S, Palacholla R, Kamarthi S, Wallace BC. Learning to identify patients at risk of uncontrolled hypertension using electronic health records data. AMIA Summits on Translational Science Proceedings. 2019 2019;2019:533.
69. Ye X, Zeng QT, Facelli JC, Brixner DI, Conway M, Bray BE. Predicting optimal hypertension treatment pathways using recurrent neural networks. International journal of medical informatics. 2020 2020;139:104122.
70. Lim B. Forecasting treatment responses over time using recurrent marginal structural networks. Advances in neural information processing systems. 2018 2018;31.
71. Lee C, Yoon J, Van Der Schaar M. Dynamic-deephit: A deep learning approach for dynamic survival analysis with competing risks based on longitudinal data. IEEE Transactions on Biomedical Engineering. 2019 2019;67(1):122-33.
72. Kwon BC, Choi M-J, Kim JT, Choi E, Kim YB, Kwon S, et al. Retainvis: Visual analytics with interpretable and interactive recurrent neural networks on electronic medical records. IEEE transactions on visualization and computer graphics. 2018 2018;25(1):299-309.
73. Choi E, Bahadori MT, Sun J, Kulas J, Schuetz A, Stewart W. Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. Advances in neural information processing systems. 2016 2016;29.
74. Pham T, Tran T, Phung D, Venkatesh S. Predicting healthcare trajectories from medical records: A deep learning approach. Journal of biomedical informatics. 2017 2017;69:218-29.
75. Kraljevic Z, Bean D, Shek A, Bendayan R, Hemingway H, Yeung JA, et al. Foresight—a generative pretrained transformer for modelling of patient timelines using electronic health records: a retrospective modelling study. The Lancet Digital Health. 2024 2024;6(4):e281-e90.
76. Xu Y, Xu S, Ramprassad M, Tumanov A, Zhang C. TransEHR: Self-Supervised Transformer for Clinical Time Series Data. Machine Learning for Health (ML4H); 20232023. p. 623-35.
77. Harerimana G, Kim JW, Jang B. A multi-headed transformer approach for predicting the patient's clinical time-series variables from charted vital signs. IEEE Access. 2022 2022;10:105993-6004.

78. Sheikhalishahi S, Bhattacharyya A, Celi LA, Osmani V. An interpretable deep learning model for time-series electronic health records: Case study of delirium prediction in critical care. Artificial Intelligence in Medicine. 2023 2023;144:102659.
79. Antikainen E, Linnosmaa J, Umer A, Oksala N, Eskola M, van Gils M, et al. Transformers for cardiac patient mortality risk prediction from heterogeneous electronic health records. Scientific Reports. 2023 2023;13(1):3517.
80. Li Y, Rao S, Solares JRA, Hassaine A, Ramakrishnan R, Canoy D, et al. BEHRT: transformer for electronic health records. Scientific reports. 2020 2020;10(1):7155.
81. Labach A, Pokhrel A, Huang XS, Zuberi S, Yi SE, Volkovs M, et al. Duett: Dual event time transformer for electronic health records. Machine Learning for Healthcare Conference; 20232023. p. 403-22.
82. Bica I, Alaa AM, Jordon J, van der Schaar M. Estimating counterfactual treatment outcomes over time through adversarially balanced representations. arXiv preprint arXiv:2002 04083. 2020 2020.
83. Berrevoets J, Curth A, Bica I, McKinney E, van der Schaar M. Disentangled counterfactual recurrent networks for treatment effect inference over time. arXiv preprint arXiv:2112 03811. 2021 2021.
84. Grecov P, Bandara K, Bergmeir C, Ackermann K, Campbell S, Scott D, et al. Causal inference using global forecasting models for counterfactual prediction. Pacific-Asia Conference on Knowledge Discovery and Data Mining; 20212021. p. 282-94.
85. Wang Z, Samsten I, Papapetrou P. Counterfactual explanations for survival prediction of cardiovascular ICU patients. Artificial Intelligence in Medicine: 19th International Conference on Artificial Intelligence in Medicine, AIME 2021, Virtual Event, June 15--18, 2021, Proceedings; 20212021. p. 338-48.
86. Li R, Hu S, Lu M, Utsumi Y, Chakraborty P, Sow DM, et al. G-net: a recurrent network approach to g-computation for counterfactual prediction under a dynamic treatment regime. Machine Learning for Health; 20212021. p. 282-99.
87. Ge Q, Hu Z, Zhang K, Li S, Lin W, Jin L, et al. Recurrent neural reinforcement learning for counterfactual evaluation of public health interventions on the spread of covid-19 in the world. medRxiv. 2020 2020.
88. Chu J, Dong W, Wang J, He K, Huang Z. Treatment effect prediction with adversarial deep learning using electronic health records. BMC Medical Informatics and Decision Making. 2020 2020;20:1-14.
89. Cerqueira V, Torgo L, Soares C. A case study comparing machine learning with statistical methods for time series forecasting: size matters. Journal of Intelligent Information Systems. 2022 2022;59(2):415-33.
90. Makridakis S, Spiliotis E, Assimakopoulos V, Semenoglou A-A, Mulder G, Nikolopoulos K. Statistical, machine learning and deep learning forecasting methods: Comparisons and ways forward. Journal of the Operational Research Society. 2023 2023;74(3):840-59.
91. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. Briefings in bioinformatics. 2018 2018;19(6):1236-46.
92. Habehh H, Gohel S. Machine learning in healthcare. Current genomics. 2021 2021;22(4):291.
93. Nwanosike EM, Conway BR, Merchant HA, Hasan SS. Potential applications and performance of machine learning techniques and algorithms in clinical practice: a systematic review. International journal of medical informatics. 2022 2022;159:104679.
94. Hashir M, Sawhney R. Towards unstructured mortality prediction with free-text clinical notes. Journal of biomedical informatics. 2020;108:103489.
95. Ehrenstein V, Kharrazi H, Lehmann H, Taylor CO. Obtaining data from electronic health records. Tools and technologies for registry interoperability, registries for evaluating patient outcomes: A user's guide, 3rd edition, Addendum 2 [Internet]: Agency for Healthcare Research and

Quality (US); 2019.
96. Ford E, Carroll JA, Smith HE, Scott D, Cassell JA. Extracting information from the text of electronic medical records to improve case detection: a systematic review. Journal of the American Medical Informatics Association. 2016;23(5):1007-15.
97. Jain R, Jangra A, Saha S, Jatowt A. A survey on medical document summarization. arXiv preprint arXiv:221201669. 2022.
98. Zhang M, Zhou G, Yu W, Huang N, Liu W. A comprehensive survey of abstractive text summarization based on deep learning. Computational intelligence and neuroscience. 2022;2022.
99. Molenaar S, Maas L, Burriel V, Dalpiaz F, Brinkkemper S, editors. Medical dialogue summarization for automated reporting in healthcare. Advanced Information Systems Engineering Workshops: CAiSE 2020 International Workshops, Grenoble, France, June 8–12, 2020, Proceedings 32; 2020: Springer.
100. Deelman E, Vahi K, Juve G, Rynge M, Callaghan S, Maechling PJ, et al. Pegasus, a workflow management system for science automation. Future Generation Computer Systems. 2015;46:17-35.
101. Lee M, Cimino J, Zhu HR, Sable C, Shanker V, Ely J, et al., editors. Beyond information retrieval—medical question answering. AMIA annual symposium proceedings; 2006: American Medical Informatics Association.
102. Terol RM, Martínez-Barco P, Palomar M. A knowledge based method for the medical question answering problem. Computers in biology and medicine. 2007;37(10):1511-21.
103. Jia J, Liu H, Gong NZ. 10 Security and Privacy Problems in Large Foundation Models. AI Embedded Assurance for Cyber Systems: Springer; 2023. p. 139-59.
104. Wornow M, Xu Y, Thapa R, Patel B, Steinberg E, Fleming S, et al. The shaky foundations of large language models and foundation models for electronic health records. npj Digital Medicine. 2023;6(1):135.
105. Elazab A, Wang C, Gardezi SJS, Bai H, Hu Q, Wang T, et al. GP-GAN: Brain tumor growth prediction using stacked 3D generative adversarial networks from longitudinal MR Images. Neural Networks. 2020 2020;132:321-32.
106. Peng L, Lin L, Lin Y, Chen Y-W, Mo Z, Vlasova RM, et al. Longitudinal prediction of infant MR images with multi-contrast perceptual adversarial learning. Frontiers in neuroscience. 2021 2021;15:653213.
107. Ning Z, Zhang Y, Pan Y, Zhong T, Liu M, Shen D. LDGAN: Longitudinal-diagnostic generative adversarial network for disease progression prediction with missing structural MRI. Machine Learning in Medical Imaging: 11th International Workshop, MLMI 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Proceedings 11; 20202020. p. 170-9.
108. Devika K, Mahapatra D, Subramanian R, Oruganti VRM. Outlier-based autism detection using longitudinal structural MRI. IEEE Access. 2022 2022;10:27794-808.
109. To M-S, Sarno IG, Chong C, Jenkinson M, Carneiro G. Self-supervised lesion change detection and localisation in longitudinal multiple sclerosis brain imaging. International Conference on Medical Image Computing and Computer-Assisted Intervention; 20212021. p. 670-80.
110. Disch N, Zimmerer D. A Holistic Vision: Modeling Patient Trajectories in Longitudinal Medical Imaging. The Symbiosis of Deep Learning and Differential Equations III; 20232023.
111. Yang F, Meng R, Cho H, Wu G, Kim WH. Disentangled sequential graph autoencoder for preclinical Alzheimer's disease characterizations from ADNI study. Medical Image Computing and Computer Assisted Intervention--MICCAI 2021: 24th International Conference, Strasbourg, France, September 27--October 1, 2021, Proceedings, Part II 24; 20212021. p. 362-72.
112. Martí-Juan G, Lorenzi M, Piella G, Initiative AsDN, Others. MC-RVAE: Multi-channel recurrent variational autoencoder for multimodal Alzheimer's disease progression modelling. Neuroimage. 2023 2023;268:119892.
113. Baur C, Denner S, Wiestler B, Navab N, Albarqouni S. Autoencoders for unsupervised anomaly segmentation in brain MR images: a comparative study. Medical Image Analysis. 2021

2021;69:101952.
114. Hong J, Kang SK, Alberts I, Lu J, Sznitman R, Lee JS, et al. Image-level trajectory inference of tau pathology using variational autoencoder for Flortaucipir PET. European journal of nuclear medicine and molecular imaging. 2022 2022;49(9):3061-72.
115. Berchuck SI, Mukherjee S, Medeiros FA. Estimating rates of progression and predicting future visual fields in glaucoma using a deep variational autoencoder. Scientific Reports. 2019 2019;9(1):18113.
116. Xiao N, Qiang Y, Zhao Z, Zhao J, Lian J. Tumour growth prediction of follow-up lung cancer via conditional recurrent variational autoencoder. IET Image Processing. 2020 2020;14(15):3975-81.
117. Kmetzsch V, Becker E, Saracino D, Rinaldi D, Camuzat A, Le Ber I, et al. Disease progression score estimation from multimodal imaging and microRNA data using supervised variational autoencoders. IEEE Journal of Biomedical and Health Informatics. 2022 2022;26(12):6024-35.
118. Sauty B, Durrleman S. Longitudinal Variational Autoencoders learn a Riemannian progression model for imaging data. GeoMedIA Workshop 2022-Geometric Deep Learning in Medical Image Analysis; 20222022.
119. Yu Y, Xie Y, Thamm T, Gong E, Ouyang J, Huang C, et al. Use of deep learning to predict final ischemic stroke lesions from initial magnetic resonance imaging. JAMA network open. 2020 2020;3(3):e200772-e.
120. Zhang L, Lu L, Wang X, Zhu RM, Bagheri M, Summers RM, et al. Spatio-temporal convolutional LSTMs for tumor growth prediction by learning 4D longitudinal patient data. IEEE transactions on medical imaging. 2019 2019;39(4):1114-26.
121. Yu X, Yang Q, Tang Y, Gao R, Bao S, Cai LY, et al. Deep conditional generative model for longitudinal single-slice abdominal computed tomography harmonization. Journal of Medical Imaging. 2024 2024;11(2):024008-.
122. Neifar N, Ben-Hamadou A, Mdhaffar A, Jmaiel M, Freisleben B. Leveraging statistical shape priors in gan-based ECG synthesis. IEEE Access. 2024 2024.
123. Hartmann KG, Schirrmeister RT, Ball T. EEG-GAN: Generative adversarial networks for electroencephalograhic (EEG) brain signals. arXiv preprint arXiv:1806 01875. 2018 2018.
124. Abdelfattah SM, Abdelrahman GM, Wang M. Augmenting the size of EEG datasets using generative adversarial networks. 2018 International joint conference on neural networks (IJCNN); 20182018. p. 1-6.
125. Panwar S, Rad P, Jung T-P, Huang Y. Modeling EEG data distribution with a Wasserstein generative adversarial network to predict RSVP events. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2020 2020;28(8):1720-30.
126. Luo Y, Lu B-L. EEG data augmentation for emotion recognition using a conditional Wasserstein GAN. 2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC); 20182018. p. 2535-8.
127. Luo T-J, Fan Y, Chen L, Guo G, Zhou C. EEG signal reconstruction using a generative adversarial network with wasserstein distance and temporal-spatial-frequency loss. Frontiers in neuroinformatics. 2020 2020;14:15.
128. Adib E, Fernandez AS, Afghah F, Prevost JJ. Synthetic ecg signal generation using probabilistic diffusion models. IEEE Access. 2023 2023.
129. Yoon JS, Zhang C, Suk H-I, Guo J, Li X. Sadm: Sequence-aware diffusion model for longitudinal medical image generation. International Conference on Information Processing in Medical Imaging; 20232023. p. 388-400.
130. Xia T, Chartsias A, Wang C, Tsaftaris SA, Initiative AsDN, Others. Learning to synthesise the ageing brain without longitudinal data. Medical Image Analysis. 2021 2021;73:102169.
131. Zhang Y, Hu X, Sapkota N, Shi Y, Chen DZ. Unsupervised feature clustering improves contrastive representation learning for medical image segmentation. 2022 IEEE International

Conference on Bioinformatics and Biomedicine (BIBM); 20222022. p. 1820-3.
132. Painchaud N, Skandarani Y, Judge T, Bernard O, Lalande A, Jodoin P-M. Cardiac segmentation with strong anatomical guarantees. IEEE transactions on medical imaging. 2020 2020;39(11):3703-13.
133. Mahapatra D, Antony B, Sedai S, Garnavi R. Deformable medical image registration using generative adversarial networks. 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018); 20182018. p. 1449-53.
134. Qiao J, Lai Q, Li Y, Lan T, Yu C, Wang X. A GAN based multi-contrast modalities medical image registration approach. 2020 IEEE International Conference on Image Processing (ICIP); 20202020. p. 3000-4.
135. Tanner C, Ozdemir F, Profanter R, Vishnevsky V, Konukoglu E, Goksel O. Generative adversarial networks for MR-CT deformable image registration. arXiv preprint arXiv:1807 07349. 2018 2018.
136. Yang H, Qian P, Fan C, Others. An indirect multimodal image registration and completion method guided by image synthesis. Computational and mathematical methods in medicine. 2020 2020;2020.
137. Kim B, Han I, Ye JC. Diffusemorph: Unsupervised deformable image registration using diffusion model. European conference on computer vision; 20222022. p. 347-64.
138. Zhang F, Wells WM, O'Donnell LJ. Deep diffusion MRI registration (DDMReg): a deep learning method for diffusion MRI registration. IEEE transactions on medical imaging. 2021 2021;41(6):1454-67.
139. Gupta S, Keshari A, Das S. Rv-gan: Recurrent gan for unconditional video generation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 20222022. p. 2024-33.
140. Celard P, Seara Vieira A, Sorribes-Fdez JM, Romero R, Lorenzo Iglesias E, Borrajo Diz L. Study on synthetic video generation of embryo development. International Conference on Hybrid Artificial Intelligence Systems; 20232023. p. 623-34.
141. Ferroukhi M, Ouahabi A, Attari M, Habchi Y, Taleb-Ahmed A. Medical video coding based on 2nd-generation wavelets: Performance evaluation. Electronics. 2019 2019;8(1):88.
142. Li C, Liu H, Liu Y, Feng BY, Li W, Liu X, et al. Endora: Video Generation Models as Endoscopy Simulators. arXiv preprint arXiv:2403 11050. 2024 2024.
143. Reynaud H, Vlontzos A, Dombrowski M, Gilligan Lee C, Beqiri A, Leeson P, et al. D'artagnan: Counterfactual video generation. International Conference on Medical Image Computing and Computer-Assisted Intervention; 20222022. p. 599-609.
144. Quellec G, Cazuguel G, Cochener B, Lamard M. Multiple-instance learning for medical image and video analysis. IEEE reviews in biomedical engineering. 2017 2017;10:213-34.
145. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. 2010 2010/1;20(1):110-21. doi: 10.1101/gr.097857.109.
146. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003 2003/7/1;31(13):3812-4. doi: 10.1093/nar/gkg509.
147. Raimondi D, Tanyalcin I, Ferté J, Gazzo A, Orlando G, Lenaerts T, et al. DEOGEN2: prediction and interactive visualization of single amino acid variant deleteriousness in human proteins. Nucleic Acids Res. 2017 2017/7/3;45(W1):W201-W6. doi: 10.1093/nar/gkx390.
148. Favalli V, Tini G, Bonetti E, Vozza G, Guida A, Gandini S, et al. Machine learning-based reclassification of germline variants of unknown significance: The RENOVO algorithm. Am J Hum Genet. 2021 2021/4/1;108(4):682-95. doi: 10.1016/j.ajhg.2021.03.010.
149. Carter H, Douville C, Stenson PD, Cooper DN, Karchin R. Identifying Mendelian disease genes with the variant effect scoring tool. BMC Genomics. 2013 2013/5/28;14 Suppl 3(Suppl 3):S3. doi: 10.1186/1471-2164-14-S3-S3.
150. Jagadeesh KA, Wenger AM, Berger MJ, Guturu H, Stenson PD, Cooper DN, et al. M-CAP

eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. Nat Genet. 2016 2016/12;48(12):1581-6. doi: 10.1038/ng.3703.
151. Pejaver V, Urresti J, Lugo-Martinez J, Pagel KA, Lin GN, Nam H-J, et al. Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. Nat Commun. 2020 2020/11/20;11(1):5918. doi: 10.1038/s41467-020-19669-x.
152. Brandes N, Goldman G, Wang CH, Ye CJ, Ntranos V. Genome-wide prediction of disease variant effects with a deep protein language model. Nat Genet. 2023 2023/9;55(9):1512-22. doi: 10.1038/s41588-023-01465-0.
153. Frazer J, Notin P, Dias M, Gomez A, Min JK, Brock K, et al. Disease variant prediction with deep generative models of evolutionary data. Nature. 2021 2021/11;599(7883):91-5. doi: 10.1038/s41586-021-04043-8.
154. Li C, Zhi D, Wang K, Liu X. MetaRNN: differentiating rare pathogenic and rare benign missense SNVs and InDels using deep learning. Genome Med. 2022 2022/10/8;14(1):115. doi: 10.1186/s13073-022-01120-z.
155. Qi H, Zhang H, Zhao Y, Chen C, Long JJ, Chung WK, et al. MVP predicts the pathogenicity of missense variants by deep learning. Nat Commun. 2021 2021/1/21;12(1):510. doi: 10.1038/s41467-020-20847-0.
156. Sundaram L, Gao H, Padigepati SR, McRae JF, Li Y, Kosmicki JA, et al. Predicting the clinical impact of human mutation with deep neural networks. Nat Genet. 2018 2018/8;50(8):1161-70. doi: 10.1038/s41588-018-0167-z.
157. Quang D, Chen Y, Xie X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. Bioinformatics. 2015 2015/3/1;31(5):761-3. doi: 10.1093/bioinformatics/btu703.
158. Huang Y, Zaas AK, Rao A, Dobigeon N, Woolf PJ, Veldman T, et al. Temporal dynamics of host molecular responses differentiate symptomatic and asymptomatic influenza a infection. PLoS Genet. 2011 Aug;7(8):e1002234. PMID: 21901105. doi: 10.1371/journal.pgen.1002234.
159. Baranzini SE, Mousavi P, Rio J, Caillier SJ, Stillman A, Villoslada P, et al. Transcription-based prediction of response to IFNbeta using supervised computational methods. PLoS Biol. 2005 Jan;3(1):e2. PMID: 15630474. doi: 10.1371/journal.pbio.0030002.
160. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol. 2014 Apr;32(4):381-6. PMID: 24658644. doi: 10.1038/nbt.2859.
161. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, et al. Reversed graph embedding resolves complex single-cell trajectories. Nat Methods. 2017 Oct;14(10):979-82. PMID: 28825705. doi: 10.1038/nmeth.4402.
162. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. Generalizing RNA velocity to transient cell states through dynamical modeling. Nat Biotechnol. 2020 Dec;38(12):1408-14. PMID: 32747759. doi: 10.1038/s41587-020-0591-3.
163. Marco E, Karp RL, Guo G, Robson P, Hart AH, Trippa L, et al. Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape. Proc Natl Acad Sci U S A. 2014 Dec 30;111(52):E5643-50. PMID: 25512504. doi: 10.1073/pnas.1408993111.
164. Wolf FA, Hamey FK, Plass M, Solana J, Dahlin JS, Gottgens B, et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Genome Biol. 2019 Mar 19;20(1):59. PMID: 30890159. doi: 10.1186/s13059-019-1663-x.
165. Sha Y, Qiu Y, Zhou P, Nie Q. Reconstructing growth and dynamic trajectories from single-cell transcriptomics data. Nature Machine Intelligence. 2024;6(1):25-39.
166. Jindal K, Adil MT, Yamaguchi N, Yang X, Wang HC, Kamimoto K, et al. Single-cell lineage capture across genomic modalities with CellTag-multi reveals fate-specific gene regulatory changes. Nat Biotechnol. 2023 Sep 25. PMID: 37749269. doi: 10.1038/s41587-023-01931-4.
167. Palii CG, Cheng Q, Gillespie MA, Shannon P, Mazurczyk M, Napolitani G, et al. Single-Cell

Proteomics Reveal that Quantitative Changes in Co-expressed Lineage-Specific Transcription Factors Determine Cell Fate. Cell Stem Cell. 2019 May 2;24(5):812-20 e5. PMID: 30880026. doi: 10.1016/j.stem.2019.02.006.
168. Sriyudthsak K, Shiraishi F, Hirai MY. Mathematical Modeling and Dynamic Simulation of Metabolic Reaction Systems Using Metabolome Time Series Data. Front Mol Biosci. 2016;3:15. PMID: 27200361. doi: 10.3389/fmolb.2016.00015.
169. Erdos B, Westerhuis JA, Adriaens ME, O'Donovan SD, Xie R, Singh-Povel CM, et al. Analysis of high-dimensional metabolomics data with complex temporal dynamics using RM-ASCA. PLoS Comput Biol. 2023 Jun;19(6):e1011221. PMID: 37352364. doi: 10.1371/journal.pcbi.1011221.
170. Ahmed KT, Sun J, Cheng S, Yong J, Zhang W. Multi-omics data integration by generative adversarial network. Bioinformatics. 2021 Dec 22;38(1):179-86. PMID: 34415323. doi: 10.1093/bioinformatics/btab608.
171. Cao ZJ, Gao G. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. Nat Biotechnol. 2022 Oct;40(10):1458-66. PMID: 35501393. doi: 10.1038/s41587-022-01284-4.
172. Ashuach T, Gabitto MI, Koodli RV, Saldi GA, Jordan MI, Yosef N. MultiVI: deep generative model for the integration of multimodal data. Nat Methods. 2023 Aug;20(8):1222-31. PMID: 37386189. doi: 10.1038/s41592-023-01909-9.
173. Yeo GHT, Saksena SD, Gifford DK. Generative modeling of single-cell time series with PRESCIENT enables prediction of cell trajectories with interventions. Nat Commun. 2021 May 28;12(1):3222. PMID: 34050150. doi: 10.1038/s41467-021-23518-w.
174. Yang Z, Wen J, Abdulkadir A, Cui Y, Erus G, Mamourian E, et al. Gene-SGAN: discovering disease subtypes with imaging and genetic signatures via multi-view weakly-supervised deep clustering. Nat Commun. 2024 Jan 8;15(1):354. PMID: 38191573. doi: 10.1038/s41467-023-44271-2.
175. Gutta C, Morhard C, Rehm M. Applying a GAN-based classifier to improve transcriptome-based prognostication in breast cancer. PLoS Comput Biol. 2023 Apr;19(4):e1011035. PMID: 37011102. doi: 10.1371/journal.pcbi.1011035.
176. Act A. Health insurance portability and accountability act of 1996. Public Law. 1996 1996;104:191.
177. Regulation GDP. General data protection regulation (GDPR). Intersoft Consulting, Accessed in October. 2018 2018.
178. Jadon A, Kumar S. Leveraging Generative AI Models for Synthetic Data Generation in Healthcare: Balancing Research and Privacy. 2023 International Conference on Smart Applications, Communications and Networking (SmartNets); 2023/7/25: IEEE; 2023. p. 1-4.
179. Albahri AS, Duhaim AM, Fadhel MA, Alnoor A, Baqer NS, Alzubaidi L, et al. A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. Inf Fusion. 2023 2023/8/1;96:156-91. doi: 10.1016/j.inffus.2023.03.008.
180. McGee RW. ChatGPT and Copyright Infringement: An Exploratory Study. 2023 2023/9/20. doi: 10.2139/ssrn.4578430.
181. ThankGod Chinonso E. The Impact of ChatGPT on Privacy and Data Protection Laws. 2023 2023/4/16. doi: 10.2139/ssrn.4574016.
182. Hacker P, Engel A, Mauer M. Regulating ChatGPT and other Large Generative AI Models. Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency; 2023/6/12. New York, NY, USA: Association for Computing Machinery; 2023. p. 1112-23.
183. Schumer C. Majority Leader Schumer Delivers Remarks To Launch SAFE Innovation Framework For Artificial Intelligence At CSIS. 2023 2023/6/21.
184. Artificial Intelligence Act: MEPs adopt landmark law. 2024 2024/3/13.
185. Cyberspace Administration of C. Interim Measures for the Administration of Generative

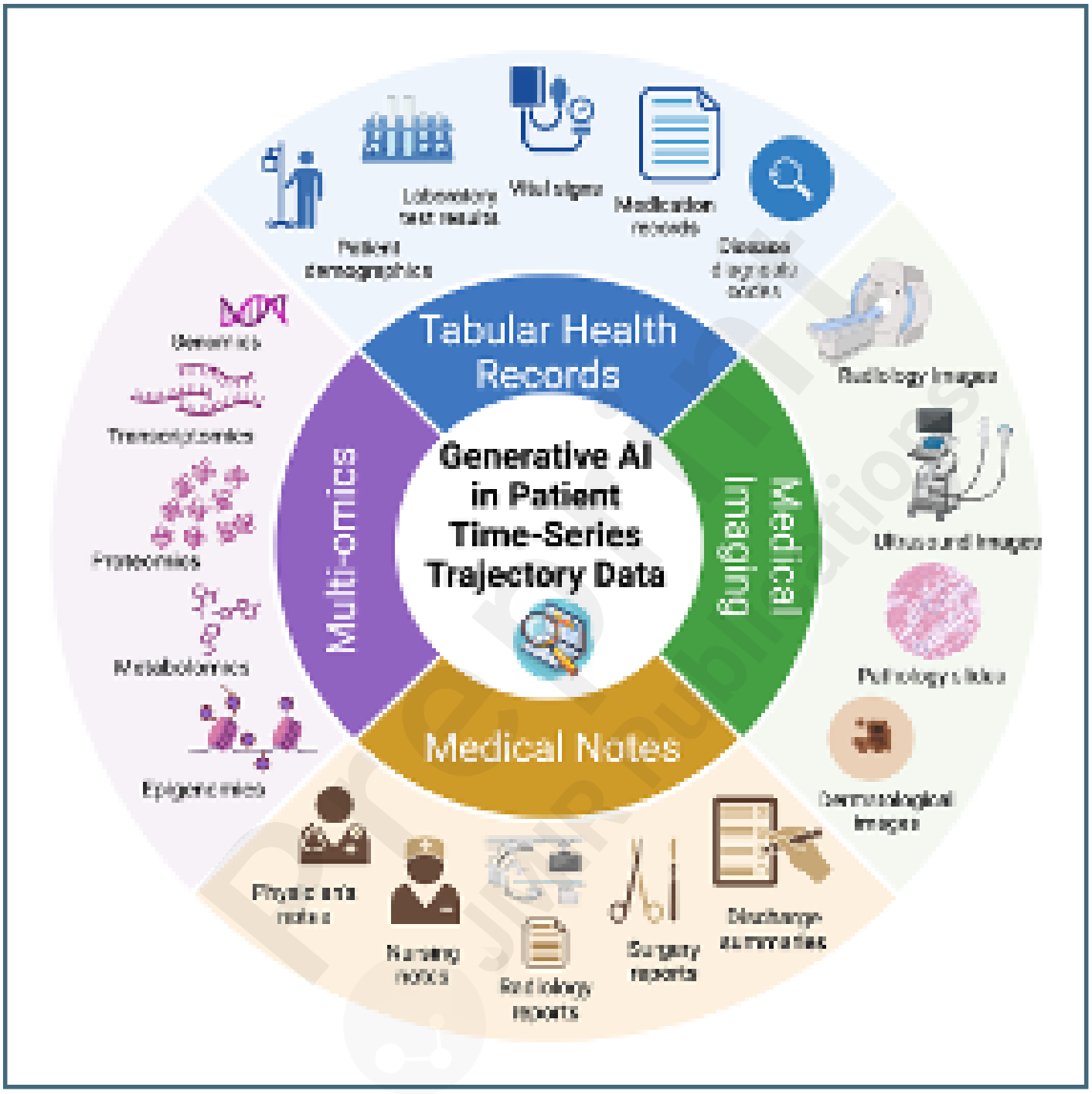
Artificial Intelligence Services. 2023 2023/7/13.
186. Miller JA, Aldosari M, Saeed F, Barna NH, Rana S, Arpinar IB, et al. A survey of deep learning and foundation models for time series forecasting. arXiv preprint arXiv:240113912. 2024.
187. Duan W, He X, Zhou L, Thiele L, Rao H, editors. Combating distribution shift for accurate time series forecasting via hypernetworks. 2022 IEEE 28th International Conference on Parallel and Distributed Systems (ICPADS); 2023: IEEE.
188. Ning Y, Teixayavong S, Shang Y, Savulescu J, Nagaraj V, Miao D, et al. Generative Artificial Intelligence in Healthcare: Ethical Considerations and Assessment Checklist. arXiv [csLG]. 2023 2023/11/2.
189. Fadler M, Legner C. Who owns data in the enterprise? Rethinking data ownership in times of big data and analytics. ECIS. 2020 2020.
190. Wang T, Zhang Y, Qi S, Zhao R, Xia Z, Weng J. Security and Privacy on Generative Data in AIGC: A Survey. arXiv [csCR]. 2023 2023/9/18.
191. Murdoch B. Privacy and artificial intelligence: challenges for protecting health information in a new era. BMC Med Ethics. 2021 2021/9/15;22(1):122. doi: 10.1186/s12910-021-00687-3.
192. Xue F, Fu Y, Zhou W, Zheng Z, You Y. To repeat or not to repeat: Insights from scaling LLM under token-crisis. Adv Neural Inf Process Syst. 2023 2023/5/22;abs/2305.13230. doi: 10.48550/arXiv.2305.13230.
193. Pakstis AJ, Speed WC, Fang R, Hyland FCL, Furtado MR, Kidd JR, et al. SNPs for a universal individual identification panel. Hum Genet. 2010 2010/3;127(3):315-24. doi: 10.1007/s00439-009-0771-1.
194. Pakstis AJ, Speed WC, Kidd JR, Kidd KK. Candidate SNPs for a universal individual identification panel. Hum Genet. 2007 2007/5;121(3-4):305-17. doi: 10.1007/s00439-007-0342-2.
195. Selvaganapathy SG, Sadasivam S. Healthcare Security: Usage of Generative Models for Malware Adversarial Attacks and Defense. Communication and Intelligent Systems; 2021: Springer Singapore; 2021. p. 885-97.
196. Notice of Data Security Incident. 2023 2023.
197. Notice of the HealthEC LLC Cyber Security Event. 2023 2023.
198. Mills J, Harclerode K. Privacy, mass intrusion and the modern data breach. Fla Law Rev. 2018 2018/3/21;69:771.
199. Greshake K, Abdelnabi S, Mishra S, Endres C, Holz T, Fritz M. Not What You've Signed Up For: Compromising Real-World LLM-Integrated Applications with Indirect Prompt Injection. Proceedings of the 16th ACM Workshop on Artificial Intelligence and Security; 2023/11/26. New York, NY, USA: Association for Computing Machinery; 2023. p. 79-90.
200. Kumar A, Agarwal C, Srinivas S, Li AJ, Feizi S, Lakkaraju H. Certifying LLM Safety against Adversarial Prompting. arXiv [csCL]. 2023 2023/9/6.
201. Salem A, Paverd A, Köpf B. Maatphor: Automated Variant Analysis for Prompt Injection Attacks. arXiv [csCR]. 2023 2023/12/12.
202. Piet J, Alrashed M, Sitawarin C, Chen S, Wei Z, Sun E, et al. Jatmo: Prompt Injection Defense by Task-Specific Finetuning. arXiv [csCR]. 2023 2023/12/29.
203. Chen S, Piet J, Sitawarin C, Wagner D. StruQ: Defending Against Prompt Injection with Structured Queries. arXiv [csCR]. 2024 2024/2/9.
204. Yu P, Xu H, Hu X, Deng C. Leveraging Generative AI and Large Language Models: A Comprehensive Roadmap for Healthcare Integration. Healthcare (Basel). 2023 2023/10/20;11(20). doi: 10.3390/healthcare11202776.
205. Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. NPJ Digit Med. 2023 2023/7/6;6(1):120. doi: 10.1038/s41746-023-00873-0.
206. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019 2019/10/25;366(6464):447-53. doi:

10.1126/science.aax2342.
207. Zack T, Lehman E, Suzgun M, Rodriguez JA, Celi LA, Gichoya J, et al. Assessing the potential of GPT-4 to perpetuate racial and gender biases in health care: a model evaluation study. Lancet Digit Health. 2024 2024/1;6(1):e12-e22. doi: 10.1016/S2589-7500(23)00225-X.
208. Bautista YJP, Theran C, Aló R, Lima V. Health Disparities Through Generative AI Models: A Comparison Study Using a Domain Specific Large Language Model. Proceedings of the Future Technologies Conference (FTC) 2023, Volume 1; 2023: Springer Nature Switzerland; 2023. p. 220-32.
209. Nazer LH, Zatarah R, Waldrip S, Ke JXC, Moukheiber M, Khanna AK, et al. Bias in artificial intelligence algorithms and recommendations for mitigation. PLOS Digit Health. 2023 2023/6;2(6):e0000278. doi: 10.1371/journal.pdig.0000278.
210. Peng C, Yang X, Chen A, Smith KE, PourNejatian N, Costa AB, et al. A study of generative large language model for medical research and healthcare. NPJ Digit Med. 2023 2023/11/16;6(1):210. doi: 10.1038/s41746-023-00958-w.
211. Wójcik MA. Foundation models in healthcare: opportunities, biases and regulatory prospects in Europe. Conference on Electronic Government and the …. 2022 2022.
212. Zhang P, Kamel Boulos MN. Generative AI in Medicine and Healthcare: Promises, Opportunities and Challenges. Future Internet. 2023 2023/8/24;15(9):286. doi: 10.3390/fi15090286.
213. Alkaissi H, McFarlane SI. Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. Cureus. 2023 2023/2;15(2):e35179. doi: 10.7759/cureus.35179.
214. Franzoni V. From Black Box to Glass Box: Advancing Transparency in Artificial Intelligence Systems for Ethical and Trustworthy AI. Computational Science and Its Applications – ICCSA 2023 Workshops; 2023: Springer Nature Switzerland; 2023. p. 118-30.
215. Sun J, Liao QV, Muller M, Agarwal M, Houde S, Talamadupula K, et al. Investigating Explainability of Generative AI for Code through Scenario-based Design. 27th International Conference on Intelligent User Interfaces; 2022/3/22. New York, NY, USA: Association for Computing Machinery; 2022. p. 212-28.
216. Dunn AG, Shih I, Ayre J, Spallek H. What generative AI means for trust in health communications. J Commun Healthc. 2023 2023/12;16(4):385-8. doi: 10.1080/17538068.2023.2277489.
217. Loh HW, Ooi CP, Seoni S, Barua PD, Molinari F, Acharya UR. Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011-2022). Comput Methods Programs Biomed. 2022 2022/11;226:107161. doi: 10.1016/j.cmpb.2022.107161.
218. Berry P, Bharadiya VM, Kumar S. Generative AI in healthcare: A trustworthy approach. 2023 2023/10/13.
219. Liu Y, Zhang K, Li Y, Yan Z, Gao C, Chen R, et al. Sora: A Review on Background, Technology, Limitations, and Opportunities of Large Vision Models. arXiv preprint arXiv:240217177. 2024.
220. Garza A, Mergenthaler-Canseco M. TimeGPT-1. arXiv preprint arXiv:231003589. 2023.
221. Yoon J, Jordon J, Van Der Schaar M. GANITE: Estimation of individualized treatment effects using generative adversarial nets. International conference on learning representations; 20182018.

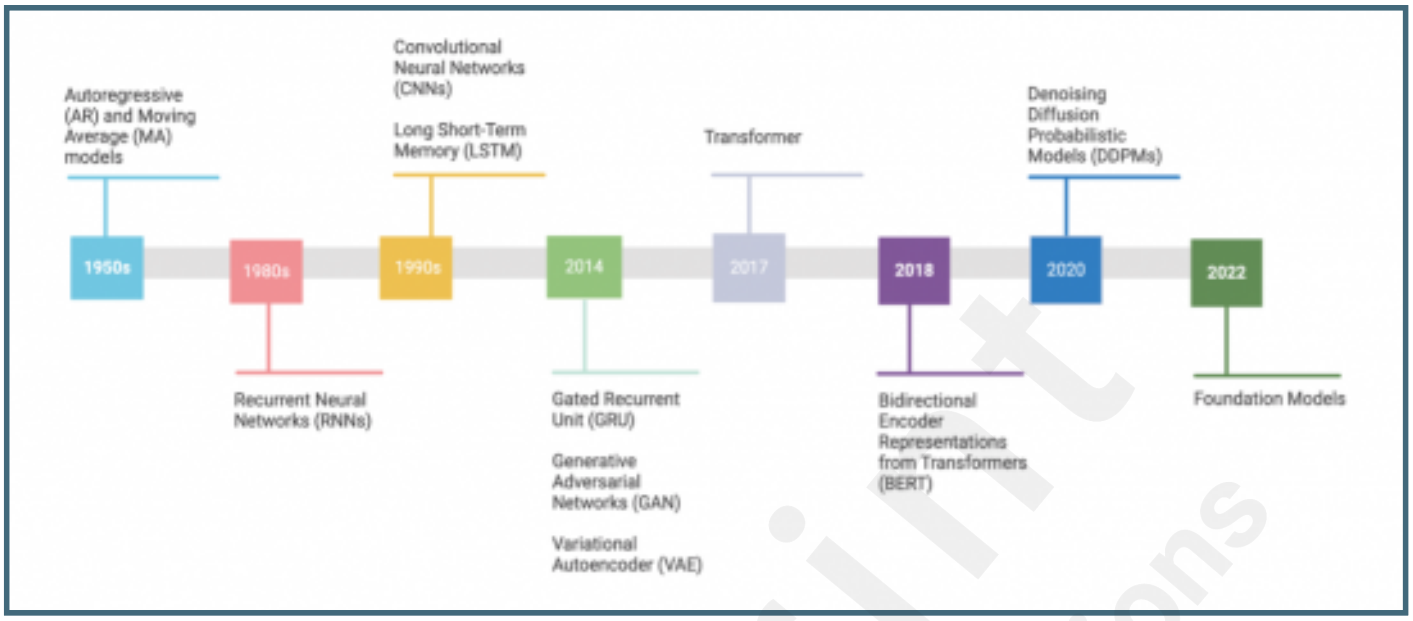
Supplementary Files

Figures

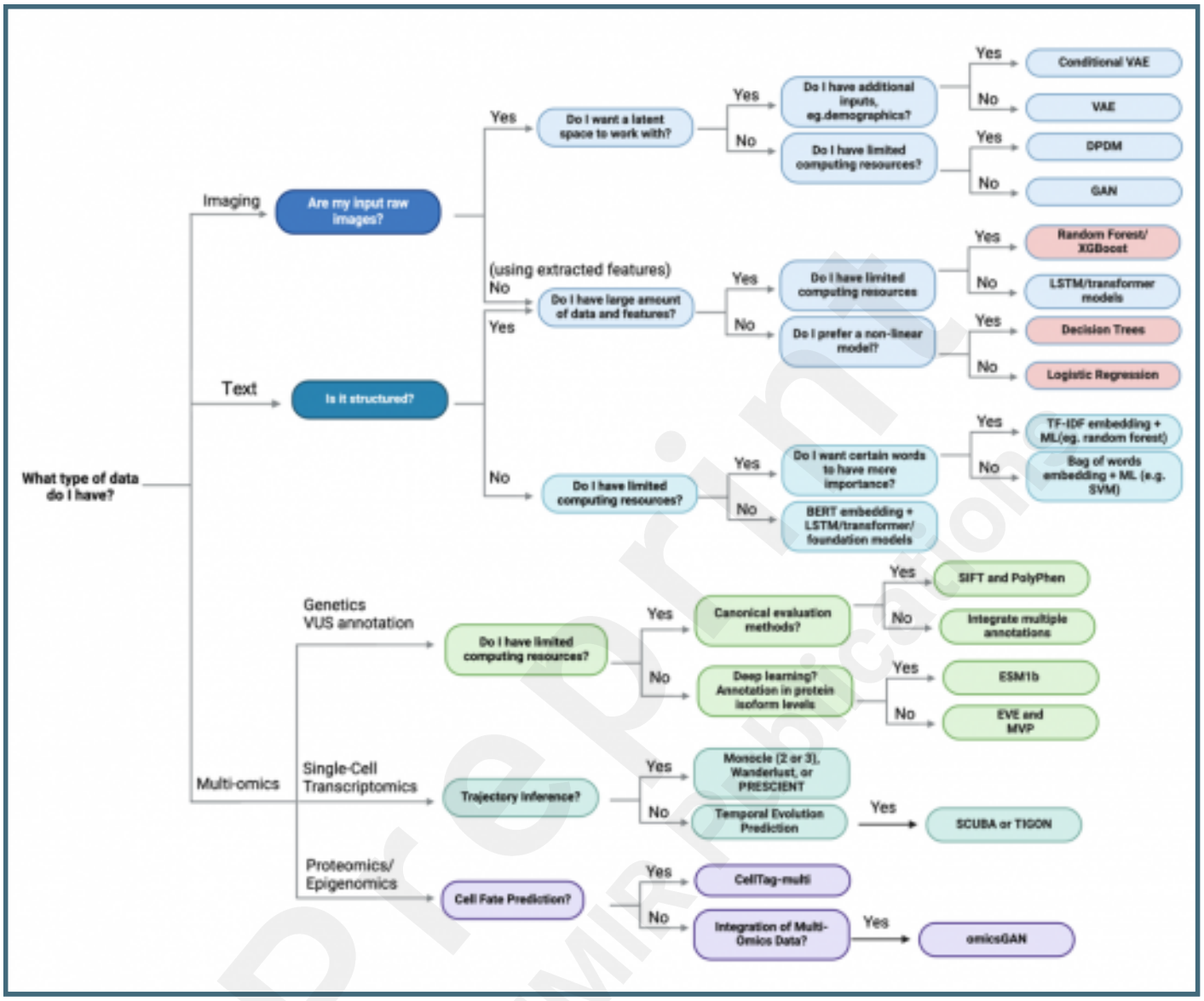
Types of data modality utilized in generative AI for time-series analysis.



Milestones in time-series forecasting models.



Workflow suggestion for choosing an appropriate model.



Multimedia Appendixes

Number of existing reviews and surveys published after 2010 from the preliminary search by year (before filtering), identified by the preliminary search with AI and biomedical related keywords: (“Generative AI” OR "Generative artificial intelligence" OR “genAI” OR “foundation model” OR “GPT” OR “Generative Adversarial Networks” OR “transformer” OR “variational autoencoder” OR “diffusion model” OR “flow model”) AND (“review” OR “summary” OR “survey”) AND (“time series”) AND (“disease trajectory” OR “healthcare”).
URL: http://asset.jmir.pub/assets/11bd4711b004bcb9c828826358a9e77f.png

Summary on existing reviews on AI in healthcare (1=limited coverage, 2=partial coverage, 3=comprehensive coverage, NA=not applicable), part 1.
URL: http://asset.jmir.pub/assets/056ad5f6066b063f5eebf4500f4b2c9c.png

Summary on existing reviews on AI in healthcare (1=limited coverage, 2=partial coverage, 3=comprehensive coverage, NA=not applicable), part 2.
URL: http://asset.jmir.pub/assets/6b6121eb7c4cbc007b916b4d17a774eb.png