UCLA Computational Medicine

Foundation Models for Physiological Signals: Opportunities and Challenges

Simon A. Lee[1] **and Kai Akamatsu**[1]
[1]UCLA Computational Medicine

This review synthesizes the paradigm shift toward foundation models for analyzing physiological signals from wearable devices. We chart the landscape from the unique characteristics of wearable data—including photoplethysmography (PPG), electrocardiography (ECG), and accelerometry (ACC)—to the architectural and pre-training innovations that enable these models. We present a comprehensive taxonomy of prominent models analyzing their contributions to tasks like health monitoring, activity recognition, and disease prediction. The core of this review is a multi-faceted examination of the critical challenges and opportunities in this emerging research field.

1. Introduction

The intersection of sensors and sophisticated machine learning has initiated a new phase in personal health monitoring. Wearable technology has transcended its status as a niche for fitness users, evolving into a mainstream platform that produces unprecedented amounts of physiological and behavioral data. This longitudinal data, collected "in the wild" during everyday activities, provides a transformative perspective for understanding human health, identifying diseases, and implementing personalized interventions. Nevertheless, the characteristics of this data (e.g. its susceptibility to artifacts, incompleteness, and variability) pose significant challenges that conventional computational techniques find difficult to address. This section lays the groundwork for understanding wearable biosignals, outlining the key data modalities and the inherent challenges that underscore the need for robust, generalizable foundation models.

The Proliferation of Wearable Sensing The adoption of wearable technology has grown exponentially, evolving from simple step counters to sophisticated health monitoring systems integrated into smartwatches, rings, and patches (Narayanswamy et al.). This proliferation has created a data ecosystem of immense scale, with predictions suggesting over 1.1 billion connected wearable devices would be in use by 2022, driving substantial cost savings in the global healthcare industry (Vijayan et al., 2021). These devices, worn on various parts of the body, are instrumental in a wide array of applications, including daily health and safety monitoring, chronic disease management, post-operative rehabilitation, athletic performance optimization, and elderly care. By enabling the continuous collection of data outside of controlled clinical settings, wearables provide a more holistic and ecologically valid picture of an individual's health status, capturing daily fluctuations and the impact of atypical events (Ferrara, 2024).

A Primer on Physiological Signal Modalities The insights from wearables are contingent on the signals they capture. While the array of sensors is ever-expanding, a few core modalities form the backbone of modern wearable health monitoring.

Photoplethysmography (PPG): This non-invasive optical technique is a cornerstone of consumer wearables, using a light source (typically an LED) and a photodetector to measure volumetric changes in blood circulation at the skin's surface (Kim and Baek, 2023). As blood pulses through the arteries, it absorbs more light, and the resulting fluctuations in reflected or transmitted light create a waveform that provides valuable information about the cardiovascular system (Castaneda et al., 2018). Wrist-worn devices commonly use green light due to its strong absorption by hemoglobin and relative robustness against motion artifacts compared to red or infrared light.

Corresponding author(s): simonlee711@g.ucla.edu

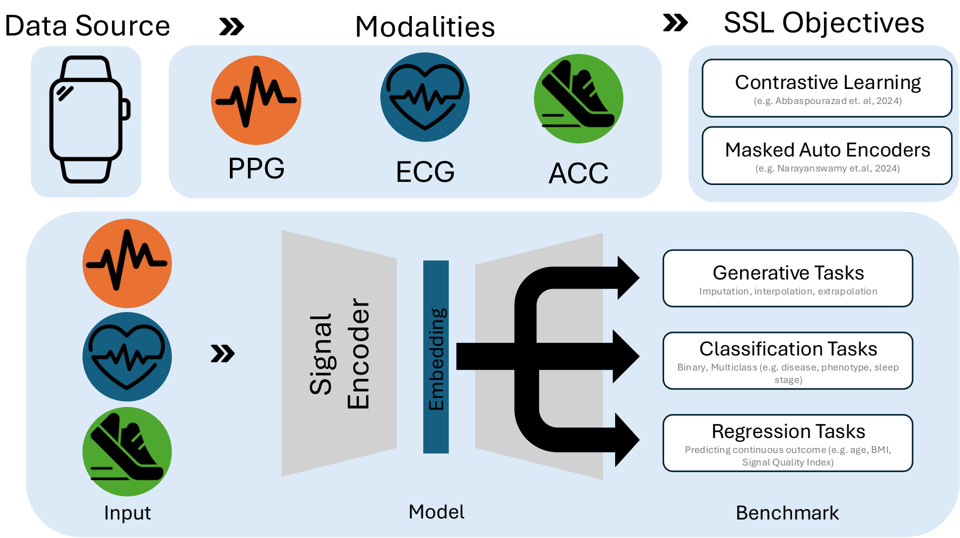

Figure 1 | **Overview of the Wearable Foundation Model Pipeline**. Data is sourced from wearable sensors and comprises multimodal physiological and motion signals including photoplethysmography (PPG), electrocardiography (ECG), and accelerometry (ACC). Self-supervised learning (SSL) objectives—such as contrastive learning and masked autoencoding, are used to pretrain a signal encoder, producing embeddings that capture rich temporal and potentially cross-modal relationships. These embeddings are then transferred to a variety of downstream processes, including generative tasks (e.g., imputation, interpolation), classification tasks (e.g., disease state, sleep stage), and regression tasks (e.g., age, BMI, signal quality index), enabling broad generalization and efficient finetuning across applications.

From this single signal, it is possible to derive critical health metrics such as heart rate (HR), heart rate variability (HRV), blood oxygen saturation (SpO2), and even infer sleep quality (Li et al., 2021; Ryals et al., 2023; Vulcan et al., 2021).

Electrocardiography (ECG): Long considered the clinical gold standard for cardiac assessment, ECG measures the heart's electrical activity (Kossmann, 1953). While traditional clinical ECGs require multiple electrodes placed across the body, the integration of single-lead ECG capabilities into consumer smart devices has been a major breakthrough (Abbaspourazad et al.). This allows for on-demand recording and the potential for detecting cardiac arrhythmias, most notably Atrial Fibrillation (AF), a leading cause of stroke. Although more convenient, single-lead ECGs are less comprehensive than their clinical counterparts, and PPG is often positioned as an even more accessible, cost-effective alternative for certain applications like HRV monitoring in healthy individuals.

Inertial Measurement Units (IMUs): IMUs are fundamental for capturing movement and are typically composed of a tri-axial accelerometer, a gyroscope, and sometimes a magnetometer (Ahmad et al., 2013; Vijayan et al., 2021). Accelerometers measure linear acceleration, gyroscopes measure angular velocity, and magnetometers measure orientation relative to the Earth's magnetic field. Together, these sensors provide a rich, multi-dimensional stream of data used for tracking physical activity, analyzing gait patterns, recognizing specific movements (e.g., running, sitting), and detecting falls—a critical application for elderly care (Narayanswamy et al.).

Other Physiological Signals: Beyond these primary modalities, many advanced wearables incorporate additional sensors. Electrodermal Activity (EDA) (Caruelle et al., 2019), also known as Galvanic Skin Response (GSR) (Sharma et al., 2016), measures changes in skin conductance related to sweat gland activity, providing a proxy for sympathetic nervous system arousal and emotional stress. Skin temperature sensors can track thermal fluctuations, and barometric altimeters can measure changes in elevation, adding context to activity data.

The Inherent Challenges of Wearables Data The immense potential of wearable data is tempered by significant, inherent challenges that arise from collecting sensitive physiological signals in uncontrolled, real-world environments.

Signal Quality and Noise: Wearable signals are notoriously susceptible to noise and artifacts. Motion Artifacts (MA) are the most pervasive challenge, particularly for PPG signals. Physical movements, such as walking or gesturing, can cause the sensor to shift relative to the skin, corrupting the optical signal and leading to inaccurate measurements of heart rate and other derived parameters (Kim and Baek, 2023). This problem is so significant that the accuracy of many PPG-based devices degrades substantially during periods of high-intensity physical activity (Mishra and Nirala, 2020).

Data Incompleteness: Unlike data collected in a lab, real-world wearable data is rarely continuous. Gaps in the data are common and can arise from numerous sources: sensor malfunctions, intermittent sensor deactivation to conserve power, poor user compliance (e.g., wearing the device too loosely), or periods when the device is removed for charging (Xu et al., 2025b). These missing data streams complicate longitudinal analysis and can render traditional time-series methods ineffective.

Heterogeneity: The wearable ecosystem is fragmented, with a vast array of devices from different manufacturers, each with its own proprietary hardware, sensor specifications, and sampling rates (de Arriba-Pérez et al., 2016). Furthermore, data varies significantly based on where the sensor is placed on the body (e.g., wrist vs. chest), the demographic characteristics of the user population (e.g., age, sex, skin tone), and the clinical or environmental context (Chen et al., 2025). This heterogeneity poses a monumental hurdle for developing universal models; an algorithm trained on data from one specific device and population may fail to generalize to another, limiting its real-world applicability (Jha et al., 2025).

The landscape of wearable signals is thus defined by a fundamental duality between signal quality and device portability. On one hand, clinical-grade instruments like multi-lead ECG systems provide high-fidelity, reliable data but are inconvenient, expensive, and confined to the clinic (Castaneda et al., 2018). On the other hand, consumer wearables prioritize convenience, comfort, and accessibility, but this comes at the cost of lower signal fidelity and greater susceptibility to noise and artifacts (Kim and Baek, 2023).

This reality has precipitated a significant philosophical shift in how data imperfections are handled. Historically, data issues like missingness and noise were treated as errors to be “fixed" during a preprocessing step, for example, through filtering or statistical imputation (Ferrara, 2024). The foundation model paradigm, however, reframes this perspective. It posits that these imperfections are not just bugs to be squashed but are themselves features to be modeled. They contain valuable contextual information about the data generation process. For instance, the fact that data is consistently missing every night between 11 PM and 7 AM is a strong indicator of the user's sleep and charging behavior. Models that learn directly from incomplete data, such as LSM-2 with its "Adaptive and Inherited Masking" technique (Xu et al., 2025b), are designed to comprehend this context. By explicitly modeling both existing ("inherited") and artificially introduced missingness, these models learn underlying physiological patterns that are robust to such gaps. This represents a move away from simple data correction toward a more sophisticated data comprehension, a necessary evolution for making sense of data from the wild.

2. Adapting Foundation Models for Physiological Signals

To address the profound challenges posed by wearable biosignal data, the research community has turned to foundation models, adapting architectures that have revolutionized other fields of AI. These models are designed to learn rich, generalizable representations from the complex temporal patterns inherent in physiological time series. This section dissects the core architectural components and design principles that are enabling this transformation, with a specific focus on their application to wearable signals.

2.1. Time Series Foundation Models (TSFMs)

Inspired by the monumental success of large pre-trained models in Natural Language Processing (NLP) (Grattafiori et al., 2024; OpenAI et al., 2024; Team et al., 2025; Zhao et al., 2023), Computer Vision (CV) (Awais et al., 2025), and even health (Wornow et al., 2023), the concept of Time-Series Foundation Models (TSFMs) has emerged as a promising new frontier (Liang et al., 2024). The central idea is to pre-train a single, large, general-purpose model on vast and diverse time-series datasets. This pre-trained model learns a fundamental "understanding" of temporal dynamics, which can then be adapted, or fine-tuned, for a wide spectrum of downstream tasks, including forecasting, classification, anomaly detection, and imputation. The core value proposition of TSFMs lies in their ability to generate powerful, task-agnostic representations directly from raw data, thereby reducing the reliance on manual feature engineering and effectively leveraging the massive quantities of unlabeled data that are characteristic of the wearable domain.

Across the landscape of TSFMs, the Transformer architecture (Vaswani et al., 2017) has unequivocally emerged as the dominant paradigm (with few emerging state space models being actively explored (Erturk et al., 2025)). Originally developed for machine translation, the Transformer's core innovation is the self-attention mechanism, which allows the model to weigh the importance of different elements in a sequence, regardless of their position. This capability is exceptionally well-suited for time-series analysis, as it enables the model to capture complex, long-range temporal dependencies and periodic patterns that are often missed by traditional recurrent or convolutional models. Models such as Lag-Llama explicitly leverage the power of Transformer architectures to perform sophisticated tasks like probabilistic time-series forecasting (Thakur, 2024).

2.2. Tokenization of Time Series Data

A fundamental challenge in applying Transformer models to time-series data is tokenization: the process of converting a continuous, analog signal into a sequence of discrete tokens that the model can process. This is a non-trivial step that has a significant impact on both model performance and computational efficiency.

A primary architectural distinction within TSFMs is the choice between patch-based and non-patch approaches (Liang et al., 2024). Treating every single time point in a long sequence (e.g., a full day of data sampled at 1 Hz contains 86,400 points) as an individual token is computationally infeasible for standard Transformers, as the complexity of the self-attention mechanism scales quadratically with the sequence length ($O(n^2)$) (Vaswani et al., 2017). To circumvent this, patching has become a popular and effective strategy. In this approach, the time series is first divided into a sequence of smaller, fixed-length, non-overlapping segments, or "patches." Each patch is then treated as a single token and is converted into a vector embedding via a linear projection. This dramatically reduces the length of the sequence fed into the Transformer, making the computation tractable.

This move towards patching reveals a deeper architectural trend: the "image-ification" of time-series data. Instead of viewing a biosignal as a one-dimensional sequence, these models effectively treat it as a two-dimensional "image," where one axis is time and the other represents different sensor channels or features. This conceptual reframing is powerful because it allows researchers to directly adapt the highly successful and heavily optimized Vision Transformer (ViT) architecture (Dosovitskiy et al., 2021), which was designed to process images by breaking them into patches (Zhang et al., 2025). While one might intuitively assume that models developed for audio signals would be a more natural fit for time-series data, the "image-ification" approach has proven to be a remarkably effective and efficient abstraction (Narayanswamy et al.). It suggests that the spatial relationships between image patches, which ViTs are designed to learn, are analogous to the temporal relationships between segments of a physiological signal.

Frequency-domain representations of time-series data also transform one-dimensional sequences into image-like forms. For instance, NormWear uses the Continuous Wavelet Transform (CWT) to decompose the signal into its time-frequency representation, creating a multi-scale representation that is compatible with physiological signals like EEG, which are characterized by diverse waveforms. This is coupled with a channel-aware attention mechanism designed specifically to handle the heterogeneity of different sensor modalities like ECG, PPG, and IMU, which have vastly different characteristics (Luo et al., 2024).

2.3. Handling Multimodality

Wearable data is inherently multimodal, comprising simultaneous streams from various sensors. An effective foundation model must be capable of processing and integrating these diverse data types, which can be either univariate (a single data stream) or multivariate (multiple concurrent streams) (Liang et al., 2024).

The most advanced models to this date are explicitly designed for multimodality, as fusing information from different sensors can lead to richer and more robust representations. The Large Sensor Model (LSM) (Narayanswamy et al.), for example, is pre-trained on a suite of signals including HR, HRV, accelerometer, EDA, and skin temperature. The architectures of these advanced multimodal models reveal a converging design pattern that can be described as a "hub-and-spoke" model. In this architecture, individual sensor streams (the "spokes") are first processed by modality-specific or channel-aware encoders. For example, SleepFM processes signals from each modality (brain activity, ECG, EMG, respiratory patterns) separately using convolutional layers before feeding them into a central transformer (Thapa et al., 2024). Similarly, NormWear processes each sensor stream individually with a shared-weight encoder before applying a channel-aware attention mechanism to integrate them (Luo et al., 2024). The outputs of these "spokes" are then fused into a central, shared representation space (the "hub"), where cross-modal relationships and dependencies are learned. This fusion can be accomplished through various techniques, such as cross-attention mechanisms, as seen in the multimodal decoder of SensorLM (Zhang et al., 2025), or through the use of a special shared classification (CLS) token, as in NormWear, which is designed to aggregate information across all channels. This modular hub-and-spoke design provides a highly scalable and flexible framework. It can gracefully handle heterogeneous sensor inputs with varying characteristics, such as channel missingness, and allows for the easy integration of new sensor modalities in the future by simply designing a new "spoke" that feeds into the central "hub."

3. Pre-training Paradigms

The defining characteristic of a foundation model is its ability to learn from vast quantities of data without requiring explicit human-provided labels. This is achieved through self-supervised learning (SSL), a paradigm that leverages the inherent structure of the data itself to create learning objectives. For wearable signals, SSL enables models to develop a fundamental understanding of physiological and behavioral patterns, which can then be transferred to a wide range of health-related tasks. This section details the key pre-training paradigms that are powering the development of foundation models for biosignals.

Self-supervised Learning: Self-supervised learning is the engine behind the foundation model. It addresses one of the greatest bottlenecks in applying deep learning to healthcare: the costly and time-consuming nature of acquiring large, expertly annotated medical datasets (Abbaspourazad et al.). In the wearable domain, unlabeled data is abundant—every second of recorded data is a potential training sample. SSL harnesses this abundance by creating "pre-text" tasks where the data itself provides the supervision (Abbaspourazad et al.). The goal is not to solve a specific downstream task during pre-training, but rather to encourage the model to learn versatile and generalizable representations of the data. These learned representations should encapsulate the complex temporal dependencies, periodicities, and cross-channel correlations present in physiological signals, making them a powerful starting point for subsequent fine-tuning on specific tasks with much smaller amounts of labeled data.

3.1. Established SSL Objectives for Wearable Signals

Several distinct families of pretext tasks have proven effective for pre-training on wearable signals.

Generative / Masked Signal Modeling: This is arguably the most prevalent SSL approach for TSFMs. Inspired by the success of Masked Autoencoders (MAEs) (He et al., 2022) in computer vision and masked language modeling in NLP, this technique involves randomly masking, or hiding, portions of the input signal and training the model to reconstruct the original, unmasked signal. To successfully reconstruct the missing data, the model cannot simply memorize the input; it must develop a deep, contextual understanding of the signal's underlying structure, such as the typical morphology of a heartbeat or the rhythmic patterns of gait. This is the primary pre-training strategy for the Large Sensor Model (LSM) (Narayanswamy et al.), which randomly masks patches across both the time and sensor axes, and for LIFT-PD, which uses a masked value prediction task to learn representations of movement data from accelerometers (Soumma et al., 2024).

Contrastive Learning: This family of methods learns representations by comparing and contrasting different samples. The core principle is to train the model to pull "positive pairs" (samples that are semantically similar) closer together in a high-dimensional embedding space, while simultaneously pushing "negative pairs" (samples that are dissimilar) far apart (Jaiswal et al., 2020). The key challenge lies in defining what constitutes a "positive pair" for unlabeled biosignal data. A highly effective strategy that has emerged is participant-level positive pairing. In this approach, different segments of a signal recorded from the same individual are treated as a positive pair, while segments from different individuals are treated as negative pairs (Abbaspourazad et al.). This pretext task implicitly forces the model to learn a unique physiological "signature" for each person, capturing the stable, idiosyncratic patterns that differentiate one individual's biosignals from another's. This is the strategy used by the Apple PPG/ECG foundation models, which defines positive pairs as augmented views of two distinct segments of data from the same participant (Abbaspourazad et al.). More advanced contrastive methods are also being developed, such as the novel "leave-one-out" contrastive learning used by SleepFM, which aligns embeddings from each modality with those of all other modalities (e.g. brain activity, ECG) from the same time window (Thapa et al., 2024). To promote discriminative and informative representations, additional regularization—such as the Kozachenko-Leonenko differential entropy estimator—can be incorporated into the loss function to encourage embeddings to be uniformly distributed in the latent space (Abbaspourazad et al.; Jing et al., 2021).

Predictive and Relational Tasks: A third category of pretext tasks involves learning temporal relationships within the signal. These tasks can include forecasting future values based on past values, predicting the correct temporal order of a set of shuffled signal segments, or predicting the time interval between two events within the signal (Ding and Wu, 2024). Some research has combined several of these objectives in a multitask SSL framework, for instance, by simultaneously training a model to recognize when a segment has been reversed, shuffled, and temporally distorted (Yuan et al., 2024a,b).

These dominant SSL paradigms reflect a fundamental duality in the learning objectives required for a comprehensive understanding of physiological data. Masked Signal Modeling, with the objective of reconstruction, compels the model to learn the universal patterns that are inherent in human physiology. To accurately reconstruct a masked portion of an ECG signal, the model must have an internal representation of a generic heartbeat. It is learning the "grammar" of physiology. In contrast, participant-level contrastive learning, with the intention of distinguishing between subjects, explicitly forces the model to learn the person-specific signatures that make an individual's physiology unique. It learns to identify the subtle variations in heart rate dynamics or gait patterns that differentiate one individual from another. A truly robust biosignal foundation model must be effective in both strategies: understanding the general rules of health time series (e.g., what constitutes a normal heart rhythm) and accounting for individual-specific baselines (e.g., what constitutes a normal rhythm for this particular person). This suggests that the future of pre-training will likely involve hybrid approaches that explicitly combine these generative and contrastive objectives within a multi-task learning framework to create richer, more personalized, and ultimately more powerful representations. However, identifying effective SSL objectives is an active area of research and efforts are not constricted to these pre-training tasks.

3.2. Alternative Learning Strategies

Beyond these core pretext tasks, more learning strategies are emerging that promise to unlock even deeper levels of understanding from wearable data.

Knowledge Distillation: This is a powerful technique where a "student" model is trained to mimic the behavior and internal representations of a more powerful "teacher" model, which is typically larger or trained using more informative data modalities (Caron et al., 2021; Gou et al., 2021; Gupta et al., 2015; Hinton et al., 2015; Tian et al., 2019). In the context of wearables, this can be used to bridge the convenience-fidelity gap by transferring knowledge from high-fidelity sensors to lower-fidelity ones. A compelling example is the development of an accelerometer-only foundation model. While an accelerometer (ACC) primarily measures motion, it indirectly captures physiological information. By training an ACC-based student model to reproduce the representations of a pre-trained and frozen PPG-based teacher model (which has direct access to cardiovascular information), the student model can learn to infer physiological states like heart rate with much greater accuracy. This approach has been shown to yield performance improvements of 23-49% for predicting HR and HRV from motion data alone, effectively creating a "virtual" PPG sensor from a simple accelerometer (Abbaspourazad et al., 2024).

In-Context Fine-Tuning: An emerging paradigm, inspired by the prompting capabilities of large language models, is in-context fine-tuning. Here, a pre-trained model is not permanently updated for a new task. Instead, at inference time, it is provided with a "prompt" consisting of several examples of related time series from the target domain. The model uses these in-context examples to adapt its predictions for the target signal on the fly, without any changes to its underlying weights. This has shown remarkable performance gains over traditional supervised methods and other foundation model approaches (Das et al., 2024).

Sensor-Language Alignment: Perhaps the most significant paradigm shift is the move towards training models not just on signals, but on pairs of signals and their corresponding natural language descriptions. Models like SensorLM (Zhang et al., 2025) are at the forefront of this trend. They use a hierarchical pipeline to automatically generate rich, descriptive captions for raw sensor data (e.g., "There is an increasing trend in heart rate between minutes 20 and 60. Biking is recording between minutes 15 and 65."). The model is then pre-trained using a hybrid objective that combines a contrastive loss (learning to match the correct textual description with a given signal segment) and a captioning loss (learning to generate the text description from the signal).

The development of these advanced strategies, particularly knowledge distillation and sensor-language alignment, reveals an implicit categorization of information content across different data modalities. Some, like accelerometer data, are information-sparse, primarily capturing raw motion. Others, like PPG and ECG, are information-rich, providing a direct window into physiological state. At the highest level, natural language provides semantic meaning and context. The most sophisticated models are no longer just fusing these signals as peers. Instead, they are using the information-rich modalities to supervise, explain, and ground the information-sparse ones. The knowledge distillation work demonstrates this clearly: PPG acts as a "teacher" to the ACC "student" (Abbaspourazad et al., 2024). SensorLM takes this a step further, using language—the ultimate semantic representation—to annotate and provide meaning to the raw numerical data from the sensors (Zhang et al., 2025). This points to a future where the central task of wearable AI is not just signal processing, but the creation of a "Rosetta Stone" for human health—a system that can seamlessly translate between the different levels of this information hierarchy, from raw data to physiological state to semantic understanding. This would enable models to reason about health in a way that is both deeply physiologically grounded and clinically meaningful.

4. Taxonomy of Foundation Models for Wearables

The theoretical advancements in architectures and pre-training paradigms have given rise to a new generation of foundation models specifically designed for wearable biosignals. These models represent the current state-of-the-art, each contributing unique innovations and targeting different facets of the wearable data challenge. This section provides a structured survey of these pioneering models, categorizing them to map the evolving landscape and highlight distinct research thrusts.

4.1. Large-Scale Multimodal Models

This category includes models that aim for broad applicability across multiple sensor types and downstream tasks, often leveraging massive, heterogeneous datasets.

LSM (Large Sensor Model) & LSM-2: The LSM family of models (Narayanswamy et al.; Xu et al., 2025b), developed by researchers at Google, represents a landmark effort in this space. Their primary contribution was to systematically demonstrate that neural scaling laws (Kaplan et al., 2020) (a principle well-established in NLP and vision) also apply to wearable sensor data. Their research empirically demonstrated that model performance on tasks like imputation and classification improves predictably as model size, dataset size, and computational budget increase. The models were pre-trained on an unprecedented dataset of 40 million hours of multimodal data (including heart rate, HRV, accelerometer, EDA, skin temperature, and altimeter) from over 165,000 individuals. The original LSM model utilized masked signal modeling as its pre-training task, where the model learned to perform random imputation, temporal interpolation, forecasting, and signal imputation (predict a subset of partially missing sensor channels). Among these, random imputation yielded the best empirical performance.

The second generation, LSM-2 (Xu et al., 2025b), introduced a critical innovation to address the pervasive issue of missing data in real-world settings. It proposed a novel self-supervised learning strategy called Adaptive and Inherited Masking (AIM), which learns robust representations directly from incomplete data without requiring a prior imputation step. AIM uses learnable mask tokens to model both the variable missingness that is inherent to the raw data ("inherited") and the missingness that is artificially introduced during training, making it significantly more robust for real-world deployment.

NormWear: NormWear (Luo et al., 2024) is a foundation model designed to handle heterogeneous sensing configurations and generalize to unseen health applications in a zero-shot manner. Its architecture incorporates several novel components. For tokenization, it uses the Continuous Wavelet Transform (CWT) to generate time-frequency representations of the input signals. Its key architectural innovation is a channel-aware attention mechanism that utilizes a shared special classification token. This mechanism allows the model to learn patterns both within a single sensor stream (intra-sensor) and between different sensor streams (inter-sensor), making it highly adaptable to various combinations of inputs (e.g., PPG, ECG, EEG, GSR, IMU). To achieve its zero-shot capabilities, NormWear is pre-trained to align the embeddings of physiological signals with the embeddings of corresponding textual descriptions, enabling it to make inferences on new health tasks defined by text prompts alone. Its generalizability has been demonstrated through extensive evaluation on 18 different applications across 11 public datasets, spanning mental health, vital sign estimation, and disease risk evaluation.

4.2. Domain- or Application-Focused Foundation Models

While generalist models aim for breadth, another research thrust focuses on developing highly optimized models for specific data modalities or clinical applications.

Domain Focused Foundation Models

Apple's PPG & ECG Foundation Models: This work is significant as it represents one of the first and largest-scale efforts to build foundation models exclusively from data collected via consumer wearable devices in a real-world setting. Using data from approximately 141,000 participants in the Apple Heart and Movement Study (AHMS), researchers trained separate models on photoplethysmography (PPG) and electrocardiogram (ECG) signals. The pre-training strategy was based on self-supervised contrastive learning, using participant-level positive pair selection to learn person-specific representations. The study demonstrated that the resulting pre-trained embeddings readily encode clinically relevant information about participants' demographics and health conditions, highlighting the potential to develop new digital biomarkers without relying on large, explicitly labeled datasets (Abbaspourazad et al.).

PaPaGei PaPaGei (Pillai et al., 2025) is an open-source foundation model tailored for photoplethysmography (PPG) signals, trained on over 57,000 hours of data encompassing 20 million unlabeled segments from public datasets. It introduces a morphology-aware contrastive learning framework that leverages domain-specific knowledge of PPG waveform structures, such as the systolic peak and dicrotic notch, to learn robust representations across diverse populations. Evaluated on 20 tasks spanning cardiovascular health, sleep disorders, pregnancy monitoring, and well-being assessment, PaPaGei outperforms existing time-series foundation models, achieving average improvements of 6.3% in classification and 2.9% in regression tasks, while being more data- and parameter-efficient. Additionally, it demonstrates robustness across different skin tones, establishing a benchmark for bias evaluation in future models.

Pulse-PPG Pulse-PPG (Saha et al., 2025) is the another open-source photoplethysmography (PPG) foundation model trained exclusively on raw, uncurated data collected over a 100-day field study involving 120 participants. Unlike prior models trained on curated clinical datasets, Pulse-PPG embraces real-world variability—motion artifacts, ambient light fluctuations, and diverse skin tones—to learn robust representations that generalize across both clinical and wearable applications. Its architecture leverages relative contrastive learning with a motif-based distance function, enabling fine-grained physiological embeddings without reliance on handcrafted features. Empirically, Pulse-PPG outperforms state-of-the-art clinical models on 10 out of 11 downstream tasks across five datasets, spanning wearable field, wearable lab, and clinical domains.

SelfPAB SelfPAB (Logacjov et al., 2024) is a self-supervised foundation model for human activity recognition (HAR) that leverages dual-accelerometer data. It is pre-trained on up to 100,000 hours of unlabeled data from the HUNT4 dataset, using a masked spectrogram reconstruction objective inspired by speech representation learning techniques. The model employs a transformer encoder architecture to learn representations that generalize across various HAR datasets, including HARTH, HAR70+, PAMAP2, Opportunity, and RealWorld. Empirical evaluations demonstrate that SelfPAB outperforms supervised baselines and other self-supervised methods, particularly in scenarios with limited labeled data, achieving F1-score improvements of up to 14%.

RelCon RelCon (Xu et al., 2025a) introduces a self-supervised learning approach for accelerometry data by leveraging a learnable motif-based distance function and a relative contrastive loss that preserves semantic similarity across time and subjects. Trained on over a billion time-series segments from 87,000+ wearable users, the resulting motion foundation model demonstrates state-of-the-art generalizability across diverse downstream tasks including gait regression and human activity recognition. The key innovation lies in modeling the degree of similarity between signals—rather than hard labels—enabling finer-grained structure in the learned representation space.

Wearable Behavioral (Feature-based) Model WBM (Wearable health Behavior Model) is a foundation model trained on over 2.5 billion hours of behavioral data from 162,000 individuals, designed to handle irregularly sampled, higher-level health signals such as activity, gait, and mobility (Erturk et al., 2025). Unlike prior work that focuses on low-level sensor streams, WBM leverages behavior-aligned timescales and physiological relevance to yield strong performance across 57 health-related tasks, including both inter-subject classification and intra-subject temporal prediction. The learned representations outperform or complement prior models like PPG-based embeddings, particularly in behavior-sensitive tasks such as sleep and pregnancy prediction, establishing behavioral modeling as an essential axis in wearable health foundation models.

Application Focused Foundation Models

Acc for Sleep Wake Logacjov et al. (2025) introduces LTA2V, a foundation model designed for sleep-wake recognition using accelerometer data. Trained on 821,700 hours of unlabeled data from the HUNT4 dataset, LTA2V employs a masked spectrogram reconstruction objective with global positional encoding to capture long-term temporal dependencies. The model demonstrates superior performance over existing self-supervised and supervised baselines, particularly in scenarios with limited labeled data, highlighting its potential for real-world applications in sleep monitoring.

SleepFM: SleepFM is a powerful multimodal foundation model designed specifically for the analysis of clinical sleep data. It was trained on a massive dataset of over 585,000 hours of polysomnography (PSG) recordings—the clinical gold standard for sleep studies—from 65,000 participants. Its architecture is channel-agnostic, meaning it can handle the variability in sensor montages commonly found in clinical sleep labs. The model's key pre-training innovation is a novel leave-one-out contrastive learning (LOO-CL) method, which learns to align the embeddings from different physiological modalities (e.g., brain activity, ECG, EMG, respiratory signals) that are recorded simultaneously during sleep. While it performs well on traditional sleep analysis tasks like sleep staging and apnea diagnosis, its most groundbreaking application is in long-term health prediction. The research demonstrated that SleepFM's learned representations of sleep physiology can predict the future onset of 130 different diseases, including dementia, Parkinson's disease, and heart failure, often years in advance (Thapa et al., 2024).

PAT (Pretrained Actigraphy Transformer): PAT is an open-source foundation model developed specifically for time-series movement data (actigraphy) from wearable accelerometers, with a focus on applications in mental health research (Ruan et al., 2025). Recognizing the computational challenge of processing very long actigraphy sequences (typically a week of data), its architecture adapts the Vision Transformer (ViT) model, using patch embeddings to efficiently encode the data. PAT was pre-trained on actigraphy data from nearly 30,000 participants using a masked autoencoder (MAE) pre-training task. When fine-tuned, it achieves state-of-the-art performance on tasks such as predicting depression, psychotropic medication status, and sleep disorders, even when the labeled fine-tuning dataset is small. A key feature of PAT is its inherent explainability; the model's attention weights can be visualized to show which periods of activity were most influential in its predictions, providing valuable insights for clinicians and researchers.

LIFT-PD (Label-efficient In-home Freezing-of-Gait Tracking): LIFT-PD is a highly specialized and efficient framework for the real-time detection of Freezing of Gait (FoG), a debilitating motor symptom in Parkinson's disease, using data from just a single accelerometer (Soumma et al., 2024). It is designed to be both label-efficient and power-efficient for practical in-home monitoring. It combines self-supervised pre-training (using a masked value prediction task) with a novel Differential Hopping Windowing Technique (DHWT) to effectively learn from imbalanced datasets where FoG events are rare. To conserve battery life on a wearable device, it incorporates an opportunistic inference module that only activates the deep learning model when active movement is detected.

5. Challenges

Despite the remarkable progress and immense promise of foundation models for wearable signals, a formidable array of challenges stands between current research prototypes. These obstacles are not isolated technical issues but form a complex, interconnected web of data-centric, algorithmic, ethical, and regulatory dilemmas that must be navigated with extreme care.

5.1. Data and Privacy Dilemma

At the very foundation of these models lies the data, and it is here that the first set of critical challenges arise.

Data Scarcity, Quality, and Heterogeneity: While raw, unlabeled wearable data is abundant, the availability of large-scale, high-quality, and expertly annotated data for specific medical conditions remains a significant bottleneck (Jha et al., 2025). Furthermore, the data that is collected in "the wild" is often plagued by artifacts, inconsistencies, and incompleteness, as discussed previously (Xu et al., 2025b). Exacerbating this is the profound challenge of data heterogeneity. Models must be able to generalize across a fragmented ecosystem of different devices, sensor types, populations, and healthcare systems (He et al., 2024). A model trained and validated on data from one population or device may see its performance plummet when deployed in a different clinical setting, a phenomenon that severely limits generalization (Lee et al., 2025b; Park et al., 2021). Accessibility is another general concern in health, and it creates a research niche that is exclusive to those with access to large scale datasets (primarily in the industry). Several works in health have explored the use of synthetic data (Gonzales et al., 2023; Lin et al., 2025) to overcome and generate "controlled" data but this has not been seen in the wearable domains yet.

Privacy and Security: Wearable data is among the most personal and sensitive information that can be collected about an individual. This creates a severe privacy and security dilemma. The large, centralized datasets required to train foundation models present an attractive target for data breaches, and there are significant risks of data misuse or unauthorized access by third parties. Ensuring strict compliance with data protection regulations such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in Europe is non-negotiable, but integrating these complex legal requirements into agile MLOps workflows is a major challenge (Venkata, 2025). To address this, researchers are exploring privacy-preserving machine learning techniques in Ai in Healthcare in general. Federated learning (FL) (Nguyen et al., 2022), for example, is a promising approach where the model is trained decentrally on data that remains on the user's local device, but this introduces its own complexities in terms of communication overhead and model performance. Differential privacy, which adds statistical noise to data to protect individual identities, is another avenue, but it often comes with a direct trade-off in model accuracy (Dwork, 2008).

5.2. Algorithmic and Computational Dilemma

The models themselves present a second set of challenges related to their scale, complexity, and transparency.

Computational Cost: Training large-scale foundation models is a computationally intensive endeavor, requiring massive clusters of specialized hardware like GPUs running for weeks or months at scale. This makes the development of such models prohibitively expensive for many academic research groups and smaller companies, potentially concentrating power in the hands of a few large technology corporations. The high energy consumption associated with this training also raises significant environmental concerns (Bender et al., 2021).

On-Device Deployment: The sheer size and computational complexity of most foundation models make it impossible to run them directly on resource-constrained wearable devices like smartwatches. This necessitates a cloud-based deployment model, where raw sensor data is transmitted from the device to a server for processing. This approach introduces latency (commonly benchmarked in speech or audio-based models (Shi et al., 2021)), which can be problematic for real-time applications, and re-introduces major privacy and security risks by moving sensitive data off-device (Li et al., 2024). Overcoming this hurdle requires significant research into model optimization techniques such as quantization (reducing the numerical precision of model weights) (Zhu et al., 2024), pruning (removing redundant connections in the network) (Cheng et al., 2024), and knowledge distillation (training a smaller model to mimic a larger one) (Abbaspourazad et al., 2024). Representations generated from these models also are prohibitive at scale and further solutions (An et al., 2025a) may be needed to run on-device inference for real time health monitoring. Innovative architectures, like the opportunistic inference module in LIFT-PD that only activates the model when necessary, also offer a path toward more efficient on-device intelligence (Soumma et al., 2024).

Interpretability and Trust (The "Black Box" Problem): A fundamental characteristic of deep neural networks, and foundation models in particular, is their opacity. They often function as "black boxes," making it exceedingly difficult to understand why a specific prediction or decision was made (Castelvecchi, 2016; Lee et al., 2025a). This lack of interpretability is a major barrier to adoption in high-stakes fields like medicine, where clinicians need to trust and be able to justify the tools they use (Jha et al., 2025). If a model provides a risk score, but cannot explain the physiological factors that contributed to it, it is unlikely to be trusted by either the doctor or the patient. Building trust requires developing robust Explainable AI (XAI) techniques (Xu et al., 2019), such as visualizing model attention weights (as done in the PAT model) or using nonlinear dynamic analysis to probe the model's internal representations (as done with NormWear) or ablating features (as done with LSM-2).

5.3. Ethical and Regulatory Considerations

Perhaps the most complex and critical challenges are those that lie at the intersection of technology, ethics, and regulation.

Algorithmic Bias: This is one of the most pressing ethical issues in AI for healthcare. Foundation models, trained on vast real-world datasets, are susceptible to learning and amplifying existing societal biases present in that data (Chen et al., 2018, 2020a,b). The most well-documented and egregious example in the wearable space is the PPG-skin tone bias. An extensive body of research has confirmed that PPG sensors that rely on green light are significantly less accurate on individuals with darker skin tones because the higher concentration of melanin in the skin absorbs more of the light, resulting in a weaker and noisier signal (Colvonen et al., 2020; Merid and Volpe, 2023; Overbye-Thompson et al., 2024). This is not a minor technical issue; it is a systemic failure that can lead to inaccurate health monitoring, false assurances, and the exacerbation of profound racial health disparities (Merid and Volpe, 2023). The sources of this bias are multi-faceted, stemming from the grassroots physics of the sensor, the lack of diversity in the datasets used for testing and validation (which are often heavily skewed towards lighter-skinned populations), and biased algorithm design.

Evaluations & Generalizability A significant portion of modern AI research is focused on achieving state of the art results, which often leads to findings that are opaque and difficult to reproduce (Arnrich et al., 2024). This issue can be attributed in part to inadequate experimental design, as well as to various malpractices in evaluation, such as selective reporting of results, lack of transparency, and the omission of critical details in published manuscripts (Arnrich et al., 2024; Lee et al., 2024). Consequently, it is essential to comprehend both the advantages and limitations of these models.

A further major challenge in the wearables, and more broadly in AI applications within healthcare, pertains to issues of generalization (Goetz et al., 2024). Given the inherent natural heterogeneity in the data (for instance, variations in device and sensor recordings, noise, etc.), there is a risk that models may not generalize effectively due to these constraints.

Clinical Validation and Regulatory Approval: There is a vast and dangerous gap between a model demonstrating high accuracy on a benchmark dataset and proving its clinical applicability (Park et al., 2021). A model can be technically proficient yet fail to improve, or even harm, patient outcomes. True clinical validation requires rigorous external testing on diverse, representative populations, ideally through prospective cohort studies or, for the highest standard of evidence, randomized controlled trials (RCTs). These studies are slow, expensive, and rarely performed before a device enters the market, leaving clinicians and patients to navigate a landscape of tools with unproven real-world benefits.

6. Future Opportunities and Research Directions

The rapid evolution of foundation models for wearable signals has opened up a new frontier in personalized health, yet the field is still in its infancy. The path forward involves not only scaling up current approaches but also fundamentally rethinking the methods we use and deepening our understanding of the data itself. Future research must bridge the gap between what is computationally possible and what is clinically meaningful, robust, and equitable.

6.1. Advancing Methodological Frontiers

The current pre-training paradigms, while effective, represent only the first wave of innovation. The next generation of models will require more sophisticated and data-aware learning objectives that move beyond the dominant generative and contrastive frameworks.

There is a significant opportunity to design novel SSL objectives tailored to the unique physics and failure modes (e.g., sensor off, artifacts) of wearable sensors (like LSM-2 (Xu et al., 2025b)). Instead of treating noise and artifacts as data to be masked or ignored, future models could be trained on pretext tasks that explicitly model these phenomena. For instance, an objective could involve learning to differentiate between signal corruption caused by motion versus that caused by poor sensor contact, thereby learning representations that are inherently robust to specific, real-world confounders. Another avenue is to leverage the natural temporal hierarchy of physiological events. A model could be trained to predict not just the next patch of a signal, but also the time until the next significant morphological feature (e.g., the next R-peak in an ECG or systolic peak in a PPG), forcing it to learn the underlying rhythm and dynamics of the signal at multiple scales.

In multimodality, the “hub-and-spoke” architecture is a promising start, but the fusion process can be much richer. The concept of an "information hierarchy," where modalities like language or high-fidelity clinical signals supervise lower-fidelity ones, points toward more advanced fusion strategies seen in this and other health domains (An et al., 2025b; Lee and Lindsey, 2024; Zhang et al., 2025). Success at this task would imply a much deeper, more integrated understanding of how behavior, motion, and physiology are linked. This could lead to a truly unified model of human health, capable of reasoning and making predictions across the full spectrum of available data streams.

6.2. Towards a Science of Wearable Data

For all the focus on model architecture, remarkably little is systematically understood about the fundamental data requirements for building robust foundation models. Key questions regarding the optimal resolution, sampling rate, duration, and diversity of training data remain largely unanswered.

There may be a critical mismatch between the sampling rates needed for optimal model performance and the rates that are practical for consumer devices due to hardware constraints. For example, while a 100Hz sampling rate might capture finer-grain details of a PPG waveform, the increased power consumption could drain a smartwatch battery far more quickly than a lower resolution alternative. Systematic studies are needed to quantify this trade-off and determine the optimal sampling rate for different downstream tasks.

Furthermore, it remains unclear which features derived from these signals are the most potent biomarkers. While we rely on established clinical metrics like Heart Rate Variability (HRV), these were developed for sparse, clinical-grade data. It's plausible that foundation models can discover novel, more sensitive digital biomarkers by learning directly from high-resolution raw signals. A crucial research direction is to develop methods that can extract and validate these learned features, translating the "black box" representations into clinically interpretable and actionable insights. The clinical information content embedded within these continuous, high-frequency signals is far from fully characterized, and a deeper collaboration between machine learning scientists and clinicians is essential to unlock their full potential.

7. Conclusion

The application of foundation models to wearable sensor data represents a significant evolution in personal health monitoring. This paper has surveyed this evolution, detailing the key architectural innovations, such as the adoption of Transformer-based models via "image-ification" of time series, and the self-supervised pre-training strategies that enable these models to learn robust physiological representations. The analysis of current models, from large-scale systems like LSM to domain-specific applications like SleepFM and LIFT-PD, illustrates a clear trend towards increasingly capable and specialized systems.

Despite this progress, formidable challenges impede widespread clinical translation. Significant obstacles include issues of data privacy, high computational cost, the "black box" problem of model interpretability, and the systemic risk of algorithmic bias. A critical gap also persists between model performance on benchmark datasets and demonstrated clinical utility, highlighting the need for more rigorous, real-world validation and stringent ethical oversight.

Future work in this domain must prioritize addressing these multifaceted challenges. Progress will depend on the development of more sophisticated SSL objectives, the establishment of a foundational science of wearable data to define optimal parameters like sampling rate and resolution, and enhanced collaboration between machine learning researchers and clinical experts. The overarching objective must extend beyond achieving incremental gains in model accuracy to focus on the creation of a robust, equitable, and trustworthy ecosystem for continuous health intelligence that can be validated in clinical practice.

Acknowledgements Simon A. Lee is supported by the Warren Alpert Fellowship. We thank the Warren Alpert Foundation for their generous support of this project, Warren Alpert-UCLA Computational Biology/AI (CBAI) Scholar Training and Retention Program (award 20242540). We declare no competing interests.

References

S. Abbaspourazad, O. Elachqar, A. Miller, S. Emrani, U. Nallasamy, and I. Shapiro. Large-scale training of foundation models for wearable biosignals. In *The Twelfth International Conference on Learning Representations*.

S. Abbaspourazad, A. Mishra, J. Futoma, A. C. Miller, and I. Shapiro. Wearable accelerometer foundation models for health via knowledge distillation. *arXiv preprint arXiv:2412.11276*, 2024.

N. Ahmad, R. A. R. Ghazilla, N. M. Khairi, and V. Kasi. Reviews on various inertial measurement unit (imu) sensor applications. *International Journal of Signal Processing Systems*, 1 (2):256–262, 2013.

U. An, M. Jeong, S. A. Lee, A. Gorla,

Y. Yang, and S. Sankararaman. Raptor: Scalable train-free embeddings for 3d medical volumes leveraging pretrained 2d foundation models. In *Forty-second International Conference on Machine Learning*, 2025a. URL https://openreview.net/forum?id=4yHWV3B6g4.

U. An, S. A. Lee, M. Jeong, A. Gorla, J. N. Chiang, and S. Sankararaman. Dk-behrt: Teaching language models international classification of disease (icd) codes using known disease descriptions. In *AAAI Bridge Program on AI for Medicine and Healthcare*, pages 133–143. PMLR, 2025b.

B. Arnrich, E. Choi, J. A. Fries, M. B. McDermott, J. Oh, T. Pollard, N. Shah, E. Steinberg, M. Wornow, and R. van de Water. Medical event data standard (meds): Facilitating machine learning for health. In *ICLR 2024 Workshop on Learning from Time Series For Health*, pages 03–08, 2024.

M. Awais, M. Naseer, S. Khan, R. M. Anwer, H. Cholakkal, M. Shah, M.-H. Yang, and F. S. Khan. Foundation models defining a new era in vision: a survey and outlook. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pages 610–623, 2021.

M. Caron, H. Touvron, I. Misra, H. Jégou, J. Mairal, P. Bojanowski, and A. Joulin. Emerging properties in self-supervised vision transformers, 2021. URL https://arxiv.org/abs/2104.14294.

D. Caruelle, A. Gustafsson, P. Shams, and L. Lervik-Olsen. The use of electrodermal activity (eda) measurement to understand consumer emotions–a literature review and a call for action. *Journal of Business Research*, 104:146–160, 2019.

D. Castaneda, A. Esparza, M. Ghamari, C. Soltanpur, and H. Nazeran. A review on wearable photoplethysmography sensors and their potential future applications in health care. *International journal of biosensors & bioelectronics*, 4(4):195, 2018.

D. Castelvecchi. Can we open the black box of ai? *Nature News*, 538(7623):20, 2016.

C.-S. Chen, Y.-J. Chen, and A. H.-W. Tsai. Large cognition model: Towards pretrained eeg foundation model. *arXiv preprint arXiv:2502.17464*, 2025.

I. Chen, F. D. Johansson, and D. Sontag. Why is my classifier discriminatory? *Advances in neural information processing systems*, 31, 2018.

I. Y. Chen, S. Joshi, and M. Ghassemi. Treating health disparities with artificial intelligence. *Nature medicine*, 26(1):16–17, 2020a.

I. Y. Chen, E. Pierson, S. Rose, S. Joshi, K. Ferryman, and M. Ghassemi. Ethical machine learning in health. *arXiv preprint arXiv:2009.10576*, 2020b.

H. Cheng, M. Zhang, and J. Q. Shi. A survey on deep neural network pruning: Taxonomy, comparison, analysis, and recommendations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

P. J. Colvonen, P. N. DeYoung, N.-O. A. Bosompra, and R. L. Owens. Limiting racial disparities and bias for wearable devices in health science research, 2020.

A. Das, M. Faw, R. Sen, and Y. Zhou. In-context fine-tuning for time-series foundation models, 2024. URL https://arxiv.org/abs/2410.24087.

F. de Arriba-Pérez, M. Caeiro-Rodríguez, and J. M. Santos-Gago. Collection and processing of data from wrist wearable devices in heterogeneous and multiple-user scenarios. *Sensors*, 16(9):1538, 2016.

C. Ding and C. Wu. Self-supervised learning for biomedical signal processing: A systematic review on ecg and ppg signals. *medRxiv*, pages 2024–09, 2024.

A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale, 2021. URL https://arxiv.org/abs/2010.11929.

C. Dwork. Differential privacy: A survey of results. In *International conference on theory and applications of models of computation*, pages 1–19. Springer, 2008.

E. Erturk, F. Kamran, S. Abbaspourazad, S. Jewell, H. Sharma, Y. Li, S. Williamson, N. J. Foti, and J. Futoma. Beyond sensor data: Foundation models of behavioral data from wearables improve health predictions. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=DtVVltU1ak.

E. Ferrara. Large language models for wearable sensor-based human activity recognition, health monitoring, and behavioral modeling: a survey of early trends, datasets, and challenges. *Sensors*, 24(15):5045, 2024.

L. Goetz, N. Seedat, R. Vandersluis, and M. van der Schaar. Generalization—a key challenge for responsible ai in patient-facing clinical applications. *npj Digital Medicine*, 7(1):126, 2024.

A. Gonzales, G. Guruswamy, and S. R. Smith. Synthetic data in health care: A narrative review. *PLOS Digital Health*, 2(1):e0000082, 2023.

J. Gou, B. Yu, S. J. Maybank, and D. Tao. Knowledge distillation: A survey. *International Journal of Computer Vision*, 129(6):1789–1819, 2021.

A. Grattafiori, A. Dubey, A. Jauhri, A. Pandey, A. Kadian, A. Al-Dahle, A. Letman, A. Mathur, A. Schelten, A. Vaughan, A. Yang, A. Fan, A. Goyal, A. Hartshorn, A. Yang, A. Mitra, A. Sravankumar, A. Korenev, A. Hinsvark, A. Rao, A. Zhang, A. Rodriguez, A. Gregerson, A. Spataru, B. Roziere, B. Biron, B. Tang, B. Chern, C. Caucheteux, C. Nayak, C. Bi, C. Marra, C. McConnell, C. Keller, C. Touret, C. Wu, C. Wong, C. C. Ferrer, C. Nikolaidis, D. Allonsius, D. Song, D. Pintz, D. Livshits, D. Wyatt, D. Esiobu, D. Choudhary, D. Mahajan, D. Garcia-Olano, D. Perino, D. Hupkes, E. Lakomkin, E. AlBadawy, E. Lobanova, E. Dinan, E. M. Smith, F. Radenovic, F. Guzmán, F. Zhang, G. Synnaeve, G. Lee, G. L. Anderson, G. Thattai, G. Nail, G. Mialon, G. Pang, G. Cucurell, H. Nguyen, H. Korevaar, H. Xu, H. Touvron, I. Zarov, I. A. Ibarra, I. Kloumann, I. Misra, I. Evtimov, J. Zhang, J. Copet, J. Lee, J. Geffert, J. Vranes, J. Park, J. Mahadeokar, J. Shah, J. van der Linde, J. Billock, J. Hong, J. Lee, J. Fu, J. Chi, J. Huang, J. Liu, J. Wang, J. Yu, J. Bitton, J. Spisak, J. Park, J. Rocca, J. Johnstun, J. Saxe, J. Jia, K. V. Alwala, K. Prasad, K. Upasani, K. Plawiak, K. Li, K. Heafield, K. Stone, K. El-Arini, K. Iyer, K. Malik, K. Chiu, K. Bhalla, K. Lakhotia, L. Rantala-Yeary, L. van der Maaten, L. Chen, L. Tan, L. Jenkins, L. Martin, L. Madaan, L. Malo, L. Blecher, L. Landzaat, L. de Oliveira, M. Muzzi, M. Pasupuleti, M. Singh, M. Paluri, M. Kardas, M. Tsimpoukelli, M. Oldham, M. Rita, M. Pavlova, M. Kambadur, M. Lewis, M. Si, M. K. Singh, M. Hassan, N. Goyal, N. Torabi, N. Bashlykov, N. Bogoychev, N. Chatterji, N. Zhang, O. Duchenne, O. Çelebi, P. Alrassy, P. Zhang, P. Li, P. Vasic, P. Weng, P. Bhargava, P. Dubal, P. Krishnan, P. S. Koura, P. Xu, Q. He, Q. Dong, R. Srinivasan, R. Ganapathy, R. Calderer, R. S. Cabral, R. Stojnic, R. Raileanu, R. Maheswari, R. Girdhar, R. Patel, R. Sauvestre, R. Polidoro, R. Sumbaly, R. Taylor, R. Silva, R. Hou, R. Wang, S. Hosseini, S. Chennabasappa, S. Singh, S. Bell, S. S. Kim, S. Edunov, S. Nie, S. Narang, S. Raparthy, S. Shen, S. Wan, S. Bhosale, S. Zhang, S. Vandenhende, S. Batra, S. Whitman, S. Sootla, S. Collot, S. Gururangan, S. Borodinsky, T. Herman, T. Fowler, T. Sheasha, T. Georgiou, T. Scialom, T. Speckbacher, T. Mihaylov, T. Xiao, U. Karn, V. Goswami, V. Gupta, V. Ramanathan, V. Kerkez, V. Gonguet, V. Do, V. Vogeti, V. Albiero, V. Petrovic, W. Chu, W. Xiong, W. Fu, W. Meers, X. Martinet, X. Wang, X. Wang, X. E. Tan, X. Xia, X. Xie, X. Jia, X. Wang, Y. Goldschlag, Y. Gaur, Y. Babaei, Y. Wen, Y. Song, Y. Zhang, Y. Li, Y. Mao, Z. D. Coudert, Z. Yan, Z. Chen, Z. Papakipos, A. Singh, A. Srivastava, A. Jain, A. Kelsey, A. Shajnfeld, A. Gangidi, A. Victoria, A. Goldstand, A. Menon, A. Sharma, A. Boesenberg, A. Baevski, A. Feinstein, A. Kallet, A. Sangani, A. Teo, A. Yunus, A. Lupu, A. Alvarado, A. Caples, A. Gu, A. Ho, A. Poulton, A. Ryan, A. Ramchandani, A. Dong, A. Franco, A. Goyal, A. Saraf, A. Chowdhury, A. Gabriel, A. Bharambe, A. Eisenman, A. Yazdan, B. James, B. Maurer, B. Leonhardi, B. Huang, B. Loyd, B. D. Paola, B. Paranjape, B. Liu, B. Wu, B. Ni, B. Hancock, B. Wasti, B. Spence, B. Stojkovic, B. Gamido, B. Montalvo, C. Parker, C. Burton, C. Mejia, C. Liu, C. Wang, C. Kim, C. Zhou, C. Hu, C.-H. Chu, C. Cai, C. Tindal, C. Feichtenhofer, C. Gao, D. Civin, D. Beaty, D. Kreymer, D. Li, D. Adkins, D. Xu, D. Testuggine, D. David, D. Parikh, D. Liskovich, D. Foss, D. Wang, D. Le, D. Holland, E. Dowling, E. Jamil, E. Montgomery, E. Presani, E. Hahn, E. Wood, E.-T. Le, E. Brinkman, E. Arcaute, E. Dunbar, E. Smothers, F. Sun, F. Kreuk, F. Tian, F. Kokkinos, F. Ozgenel, F. Caggioni, F. Kanayet, F. Seide, G. M. Florez, G. Schwarz, G. Badeer, G. Swee, G. Halpern, G. Herman, G. Sizov, Guangyi, Zhang, G. Lakshminarayanan, H. Inan, H. Shojanazeri, H. Zou, H. Wang, H. Zha, H. Habeeb, H. Rudolph, H. Suk, H. Aspegren, H. Goldman, H. Zhan, I. Damlaj, I. Molybog, I. Tufanov, I. Leontiadis, I.-E. Veliche, I. Gat, J. Weissman, J. Geboski, J. Kohli, J. Lam, J. Asher, J.-B. Gaya, J. Marcus, J. Tang, J. Chan, J. Zhen, J. Reizenstein, J. Teboul, J. Zhong, J. Jin, J. Yang, J. Cummings, J. Carvill, J. Shepard, J. McPhie, J. Torres, J. Ginsburg, J. Wang, K. Wu, K. H. U, K. Saxena, K. Khandelwal, K. Zand, K. Matosich, K. Veeraraghavan, K. Michelena, K. Li, K. Jagadeesh, K. Huang, K. Chawla, K. Huang, L. Chen, L. Garg, L. A, L. Silva, L. Bell, L. Zhang, L. Guo, L. Yu, L. Moshkovich, L. Wehrstedt, M. Khabsa, M. Avalani, M. Bhatt, M. Mankus, M. Hasson, M. Lennie, M. Reso, M. Groshev, M. Naumov, M. Lathi, M. Keneally, M. Liu, M. L. Seltzer, M. Valko, M. Restrepo, M. Patel, M. Vyatskov, M. Samvelyan, M. Clark, M. Macey, M. Wang, M. J. Hermoso, M. Metanat, M. Rastegari, M. Bansal, N. Santhanam, N. Parks, N. White, N. Bawa, N. Singhal, N. Egebo, N. Usunier, N. Mehta, N. P. Laptev, N. Dong, N. Cheng, O. Chernoguz, O. Hart, O. Salpekar, O. Kalinli, P. Kent, P. Parekh, P. Saab, P. Balaji, P. Rittner, P. Bontrager, P. Roux, P. Dollar, P. Zvyagina, P. Ratanchandani, P. Yuvraj, Q. Liang, R. Alao, R. Rodriguez, R. Ayub, R. Murthy, R. Nayani, R. Mitra, R. Parthasarathy, R. Li, R. Hogan, R. Battey, R. Wang, R. Howes, R. Rinott, S. Mehta, S. Siby, S. J. Bondu, S. Datta, S. Chugh, S. Hunt, S. Dhillon, S. Sidorov, S. Pan, S. Mahajan, S. Verma, S. Yamamoto, S. Ramaswamy, S. Lindsay, S. Lindsay, S. Feng, S. Lin, S. C. Zha, S. Patil, S. Shankar, S. Zhang, S. Zhang, S. Wang, S. Agarwal, S. Sajuyigbe, S. Chintala, S. Max, S. Chen, S. Kehoe, S. Satterfield, S. Govindaprasad, S. Gupta, S. Deng, S. Cho, S. Virk, S. Subramanian, S. Choudhury, S. Goldman, T. Remez, T. Glaser, T. Best, T. Koehler, T. Robinson, T. Li, T. Zhang, T. Matthews, T. Chou, T. Shaked, V. Vontimitta, V. Ajayi, V. Montanez, V. Mohan, V. S. Kumar, V. Mangla, V. Ionescu, V. Poenaru, V. T. Mihailescu, V. Ivanov, W. Li, W. Wang, W. Jiang, W. Bouaziz, W. Constable, X. Tang, X. Wu, X. Wang, X. Wu, X. Gao, Y. Kleinman, Y. Chen, Y. Hu, Y. Jia, Y. Qi, Y. Li, Y. Zhang, Y. Zhang, Y. Adi, Y. Nam, Yu, Wang, Y. Zhao, Y. Hao, Y. Qian, Y. Li, Y. He, Z. Rait, Z. DeVito, Z. Rosnbrick, Z. Wen, Z. Yang, Z. Zhao, and Z. Ma. The llama 3 herd of models, 2024. URL https://arxiv.org/abs/2407.21783.

S. Gupta, J. Hoffman, and J. Malik. Cross modal distillation for supervision transfer. *arXiv*, 2015. doi: 10.48550/arxiv.1507.00448. URL https://arxiv.org/abs/1507.00448.

K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16000–16009, 2022.

Y. He, F. Huang, X. Jiang, Y. Nie, M. Wang, J. Wang, and H. Chen. Foundation model for advancing healthcare: challenges, opportunities and future directions. *IEEE Reviews in Biomedical Engineering*, 2024.

G. Hinton, O. Vinyals, and J. Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

A. Jaiswal, A. R. Babu, M. Z. Zadeh, D. Banerjee, and F. Makedon. A survey on contrastive self-supervised learning. *Technologies*, 9(1):2, 2020.

D. Jha, G. Durak, A. Das, J. Sanjotra, O. Susladkar, S. Sarkar, A. Rauniyar, N. Kumar Tomar, L. Peng, S. Li, et al. Ethical framework for responsible foundational models in medical imaging. *Frontiers in Medicine*, 12:1544501, 2025.

L. Jing, P. Vincent, Y. LeCun, and Y. Tian. Understanding dimensional collapse in contrastive self-supervised learning. *arXiv*, 2021. doi: 10.48550/arxiv.2110.09348. URL https://arxiv.org/abs/2110.09348.

J. Kaplan, S. McCandlish, T. Henighan, T. B. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu, and D. Amodei. Scaling laws for neural language models, 2020. URL https://arxiv.org/abs/2001.08361.

K. B. Kim and H. J. Baek. Photoplethysmography in wearable devices: a comprehensive review of technological advances, current challenges, and future directions. *Electronics*, 12(13):2923, 2023.

C. E. Kossmann. The normal electrocardiogram. *Circulation*, 8(6):920–936, 1953.

S. A. Lee and T. Lindsey. Can large language models abstract medical coded language? *arXiv preprint arXiv:2403.10822*, 2024.

S. A. Lee, J. Lee, and J. N. Chiang. Feet: A framework for evaluating embedding techniques. *arXiv preprint arXiv:2411.01322*, 2024.

S. A. Lee, H. Halperin, Y. Halperin, T. Brokowski, and J. N. Chiang. Using foundation models to prescribe patients proper antibiotics. In J. Wu, J. Zhu, M. Xu, and Y. Jin, editors, *Proceedings of The First AAAI Bridge Program on AI for Medicine and Healthcare*, volume 281 of *Proceedings of Machine Learning Research*, pages 121–132. PMLR, 25 Feb 2025a. URL https://proceedings.mlr.press/v281/lee25a.html.

S. A. Lee, S. Jain, A. Chen, K. Ono, A. Biswas, Á. Rudas, J. Fang, and J. N. Chiang. Clinical decision support using pseudo-notes from multiple streams of ehr data. *npj Digital Medicine*, 8(1):394, 2025b.

Q. Li, Q. Li, A. S. Cakmak, G. Da Poian, D. L. Bliwise, V. Vaccarino, A. J. Shah, and G. D. Clifford. Transfer learning from ecg to ppg for improved sleep staging from wrist-worn wearables. *Physiological measurement*, 42(4):044004, 2021.

X. Li, L. Peng, Y. Wang, and W. Zhang. Open challenges and opportunities in federated foundation models towards biomedical healthcare, 2024. URL https://arxiv.org/abs/2405.06784.

Y. Liang, H. Wen, Y. Nie, Y. Jiang, M. Jin, D. Song, S. Pan, and Q. Wen. Foundation models for time series analysis: A tutorial and survey. In *Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining*, pages 6555–6565, 2024.

Y. Lin, Z. B. Yu, and S. Lee. A case study exploring the current landscape of synthetic medical record generation with commercial llms. *arXiv preprint arXiv:2504.14657*, 2025.

A. Logacjov, S. Herland, A. Ustad, and K. Bach. Selfpab: large-scale pre-training on accelerometer data for human activity recognition. *Applied Intelligence*, 54(6):4545–4563, 2024.

A. Logacjov, K. Bach, and P. J. Mork. Long-term self-supervised learning for accelerometer-based sleep–wake recognition. *Engineering Applications of Artificial Intelligence*, 141:109758, 2025.

Y. Luo, Y. Chen, A. Salekin, and T. Rahman. Toward foundation model for multivariate wearable sensing of physiological signals. *arXiv preprint arXiv:2412.09758*, 2024.

B. Merid and V. Volpe. Race correction and algorithmic bias in atrial fibrillation wearable technologies. *Health Equity*, 7(1):817–824, 2023.

B. Mishra and N. S. Nirala. A survey on denoising techniques of ppg signal. In *2020 IEEE international conference for innovation in technology (INOCON)*, pages 1–8. IEEE, 2020.

G. Narayanswamy, X. Liu, K. Ayush, Y. Yang, X. Xu, J. Garrison, S. A. Tailor, J. Sunshine, Y. Liu, T. Althoff, et al. Scaling wearable foundation models. In *The Thirteenth International Conference on Learning Representations*.

D. C. Nguyen, Q.-V. Pham, P. N. Pathirana, M. Ding, A. Seneviratne, Z. Lin, O. Dobre, and W.-J. Hwang. Federated learning for smart healthcare: A survey. *ACM Computing Surveys (Csur)*, 55(3):1–37, 2022.

OpenAI, J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat, R. Avila, I. Babuschkin, S. Balaji, V. Balcom, P. Baltescu, H. Bao, M. Bavarian, J. Belgum, I. Bello, J. Berdine, G. Bernadett-Shapiro, C. Berner, L. Bogdonoff, O. Boiko, M. Boyd, A.-L. Brakman, G. Brockman, T. Brooks, M. Brundage, K. Button, T. Cai, R. Campbell, A. Cann, B. Carey, C. Carlson, R. Carmichael, B. Chan, C. Chang, F. Chantzis, D. Chen, S. Chen, R. Chen, J. Chen, M. Chen, B. Chess, C. Cho, C. Chu, H. W. Chung, D. Cummings, J. Currier, Y. Dai, C. Decareaux, T. Degry, N. Deutsch, D. Deville, A. Dhar, D. Dohan, S. Dowling, S. Dunning, A. Ecoffet, A. Eleti, T. Eloundou, D. Farhi, L. Fedus, N. Felix, S. P. Fishman, J. Forte, I. Fulford, L. Gao, E. Georges, C. Gibson, V. Goel, T. Gogineni, G. Goh, R. Gontijo-Lopes, J. Gordon, M. Grafstein, S. Gray, R. Greene, J. Gross, S. S. Gu, Y. Guo, C. Hallacy, J. Han, J. Harris, Y. He, M. Heaton, J. Heidecke, C. Hesse, A. Hickey, W. Hickey, P. Hoeschele, B. Houghton, K. Hsu, S. Hu, X. Hu, J. Huizinga, S. Jain, S. Jain, J. Jang, A. Jiang, R. Jiang, H. Jin, D. Jin, S. Jomoto, B. Jonn, H. Jun, T. Kaftan, Łukasz Kaiser, A. Kamali, I. Kanitscheider, N. S. Keskar, T. Khan, L. Kilpatrick, J. W. Kim, C. Kim, Y. Kim, J. H. Kirchner, J. Kiros, M. Knight, D. Kokotajlo, Łukasz Kondraciuk, A. Kondrich, A. Konstantinidis, K. Kosic, G. Krueger, V. Kuo, M. Lampe, I. Lan, T. Lee, J. Leike, J. Leung, D. Levy, C. M. Li, R. Lim, M. Lin, S. Lin, M. Litwin, T. Lopez, R. Lowe, P. Lue, A. Makanju, K. Malfacini, S. Manning, T. Markov, Y. Markovski, B. Martin, K. Mayer, A. Mayne, B. McGrew, S. M. McKinney, C. McLeavey, P. McMillan, J. McNeil, D. Medina, A. Mehta, J. Menick, L. Metz, A. Mishchenko, P. Mishkin, V. Monaco, E. Morikawa, D. Mossing, T. Mu, M. Murati, O. Murk, D. Mély, A. Nair, R. Nakano, R. Nayak, A. Neelakantan, R. Ngo, H. Noh, L. Ouyang, C. O'Keefe, J. Pachocki, A. Paino, J. Palermo, A. Pantuliano, G. Parascandolo, J. Parish, E. Parparita, A. Passos, M. Pavlov, A. Peng, A. Perelman, F. de Avila Belbute Peres, M. Petrov, H. P. de Oliveira Pinto, Michael, Pokorny, M. Pokrass, V. H. Pong, T. Powell, A. Power, B. Power, E. Proehl, R. Puri, A. Radford, J. Rae, A. Ramesh, C. Raymond, F. Real, K. Rimbach, C. Ross, B. Rotsted, H. Roussez, N. Ryder, M. Saltarelli, T. Sanders, S. Santurkar, G. Sastry, H. Schmidt, D. Schnurr, J. Schulman, D. Selsam, K. Sheppard, T. Sherbakov, J. Shieh, S. Shoker, P. Shyam, S. Sidor, E. Sigler, M. Simens, J. Sitkin, K. Slama, I. Sohl, B. Sokolowsky, Y. Song, N. Staudacher, F. P. Such, N. Summers, I. Sutskever, J. Tang, N. Tezak, M. B. Thompson, P. Tillet, A. Tootoonchian, E. Tseng, P. Tuggle, N. Turley, J. Tworek, J. F. C. Uribe, A. Vallone, A. Vijayvergiya, C. Voss, C. Wainwright, J. J. Wang, A. Wang, B. Wang, J. Ward, J. Wei, C. Weinmann, A. Welihinda, P. Welinder, J. Weng, L. Weng, M. Wiethoff, D. Willner, C. Winter, S. Wolrich, H. Wong, L. Workman, S. Wu, J. Wu, M. Wu, K. Xiao, T. Xu, S. Yoo, K. Yu, Q. Yuan, W. Zaremba, R. Zellers, C. Zhang, M. Zhang, S. Zhao, T. Zheng, J. Zhuang, W. Zhuk, and B. Zoph. Gpt-4 technical report, 2024. URL https://arxiv.org/abs/2303.08774.

H. Overbye-Thompson, K. A. Hamilton, and D. Mastro. Reinvention mediates impacts of skin tone bias in algorithms: implications for technology diffusion. *Journal of Computer-Mediated Communication*, 29(5): zmae016, 2024.

S. H. Park, J. Choi, and J.-S. Byeon. Key principles of clinical validation, device approval, and insurance coverage decisions of artificial intelligence. *Korean journal of radiology*, 22(3):442, 2021.

A. Pillai, D. Spathis, F. Kawsar, and M. Malekzadeh. Papagei: Open foundation models for optical physiological signals, 2025. URL https://arxiv.org/abs/2410.20542.

F. Y. Ruan, A. Zhang, J. Y. Oh, S. Jin, and N. C. Jacobson. Ai foundation models for wearable movement data in mental health research. *ArXiv*, pages arXiv–2411, 2025.

S. Ryals, A. Chiang, S. Schutte-Rodin, A. Chandrakantan, N. Verma, S. Holfinger, F. Abbasi-Feinberg, A. Bandyopadhyay, K. Baron, S. Bhargava, et al. Photoplethysmography—new applications for an old technology: a sleep technology review. *Journal of Clinical Sleep Medicine*, 19(1):189–195, 2023.

M. Saha, M. A. Xu, W. Mao, S. Neupane, J. M. Rehg, and S. Kumar. Pulse-ppg: An open-source field-trained ppg foundation model for wearable applications across lab and field settings. *arXiv preprint arXiv:2502.01108*, 2025.

M. Sharma, S. Kacker, and M. Sharma. A brief introduction and review on galvanic skin response. *Int. J. Med. Res. Prof*, 2(6):13–17, 2016.

Y. Shi, Y. Wang, C. Wu, C.-F. Yeh, J. Chan, F. Zhang, D. Le, and M. Seltzer. Emformer: Efficient memory transformer based acoustic model for low latency streaming speech recognition. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6783–6787. IEEE, 2021.

S. B. Soumma, K. Mangipudi, D. Peterson, S. Mehta, and H. Ghasemzadeh. Wearable-based real-time freezing of gait detection in parkinson's disease using self-supervised learning. *arXiv preprint arXiv:2410.20715*, 2024.

G. Team, R. Anil, S. Borgeaud, J.-B. Alayrac, J. Yu, R. Soricut, J. Schalkwyk, A. M. Dai, A. Hauth, K. Millican, D. Silver, M. Johnson, I. Antonoglou, J. Schrittwieser, A. Glaese, J. Chen, E. Pitler, T. Lillicrap, A. Lazaridou, O. Firat, J. Molloy, M. Isard, P. R. Barham, T. Hennigan, B. Lee, F. Viola, M. Reynolds, Y. Xu, R. Doherty, E. Collins, C. Meyer, E. Rutherford, E. Moreira, K. Ayoub, M. Goel, J. Krawczyk, C. Du, E. Chi, H.-T. Cheng, E. Ni, P. Shah, P. Kane, B. Chan, M. Faruqui, A. Severyn, H. Lin, Y. Li, Y. Cheng, A. Ittycheriah, M. Mahdieh, M. Chen, P. Sun, D. Tran, S. Bagri, B. Lakshminarayanan, J. Liu, A. Orban, F. Güra, H. Zhou, X. Song, A. Boffy, H. Ganapathy, S. Zheng, H. Choe, Ágoston Weisz, T. Zhu, Y. Lu, S. Gopal, J. Kahn, M. Kula, J. Pitman, R. Shah, E. Taropa, M. A. Merey, M. Baeuml, Z. Chen, L. E. Shafey, Y. Zhang, O. Sercinoglu, G. Tucker, E. Piqueras, M. Krikun, I. Barr, N. Savinov, I. Danihelka, B. Roelofs, A. White, A. Andreassen, T. von Glehn, L. Yagati, M. Kazemi, L. Gonzalez, M. Khalman, J. Sygnowski, A. Frechette, C. Smith, L. Culp, L. Proleev, Y. Luan, X. Chen, J. Lottes, N. Schucher, F. Lebron, A. Rrustemi, N. Clay, P. Crone, T. Kocisky, J. Zhao, B. Perz, D. Yu, H. Howard, A. Bloniarz, J. W. Rae, H. Lu, L. Sifre, M. Maggioni, F. Alcober, D. Garrette, M. Barnes, S. Thakoor, J. Austin, G. Barth-Maron, W. Wong, R. Joshi, R. Chaabouni, D. Fatiha, A. Ahuja, G. S. Tomar, E. Senter, M. Chadwick, I. Kornakov, N. Attaluri, I. Iturrate, R. Liu, Y. Li, S. Cogan, J. Chen, C. Jia, C. Gu, Q. Zhang, J. Grimstad, A. J. Hartman, X. Garcia, T. S. Pillai, J. Devlin, M. Laskin, D. de Las Casas, D. Valter, C. Tao, L. Blanco, A. P. Badia, D. Reitter, M. Chen, J. Brennan, C. Rivera, S. Brin, S. Iqbal, G. Surita, J. Labanowski, A. Rao, S. Winkler, E. Parisotto, Y. Gu, K. Olszewska, R. Addanki, A. Miech, A. Louis, D. Teplyashin, G. Brown, E. Catt, J. Balaguer, J. Xiang, P. Wang, Z. Ashwood, A. Briukhov, A. Webson, S. Ganapathy, S. Sanghavi, A. Kannan, M.-W. Chang, A. Stjerngren, J. Djolonga, Y. Sun, A. Bapna, M. Aitchison, P. Pejman, H. Michalewski, T. Yu, C. Wang, J. Love, J. Ahn, D. Bloxwich, K. Han, P. Humphreys, T. Sellam, J. Bradbury, V. Godbole, S. Samangooei, B. Damoc, A. Kaskasoli, S. M. R. Arnold, V. Vasudevan, S. Agrawal, J. Riesa, D. Lepikhin, R. Tanburn, S. Srinivasan, H. Lim, S. Hodkinson, P. Shyam, J. Ferret, S. Hand, A. Garg, T. L. Paine, J. Li, Y. Li, M. Giang, A. Neitz, Z. Abbas, S. York, M. Reid, E. Cole, A. Chowdhery, D. Das, D. Rogozińska, V. Nikolaev, P. Sprechmann, Z. Nado, L. Zilka, F. Prost, L. He, M. Monteiro, G. Mishra, C. Welty, J. Newlan, D. Jia, M. Allamanis, C. H. Hu, R. de Liedekerke, J. Gilmer, C. Saroufim,

S. Rijhwani, S. Hou, D. Shrivastava, A. Baddepudi, A. Goldin, A. Ozturel, A. Cassirer, Y. Xu, D. Sohn, D. Sachan, R. K. Amplayo, C. Swanson, D. Petrova, S. Narayan, A. Guez, S. Brahma, J. Landon, M. Patel, R. Zhao, K. Villela, L. Wang, W. Jia, M. Rahtz, M. Giménez, L. Yeung, J. Keeling, P. Georgiev, D. Mincu, B. Wu, S. Haykal, R. Saputro, K. Vodrahalli, J. Qin, Z. Cankara, A. Sharma, N. Fernando, W. Hawkins, B. Neyshabur, S. Kim, A. Hutter, P. Agrawal, A. Castro-Ros, G. van den Driessche, T. Wang, F. Yang, S. yiin Chang, P. Komarek, R. McIlroy, M. Lučić, G. Zhang, W. Farhan, M. Sharman, P. Natsev, P. Michel, Y. Bansal, S. Qiao, K. Cao, S. Shakeri, C. Butterfield, J. Chung, P. K. Rubenstein, S. Agrawal, A. Mensch, K. Soparkar, K. Lenc, T. Chung, A. Pope, L. Maggiore, J. Kay, P. Jhakra, S. Wang, J. Maynez, M. Phuong, T. Tobin, A. Tacchetti, M. Trebacz, K. Robinson, Y. Katariya, S. Riedel, P. Bailey, K. Xiao, N. Ghelani, L. Aroyo, A. Slone, N. Houlsby, X. Xiong, Z. Yang, E. Gribovskaya, J. Adler, M. Wirth, L. Lee, M. Li, T. Kagohara, J. Pavagadhi, S. Bridgers, A. Bortsova, S. Ghemawat, Z. Ahmed, T. Liu, R. Powell, V. Bolina, M. Iinuma, P. Zablotskaia, J. Besley, D.-W. Chung, T. Dozat, R. Comanescu, X. Si, J. Greer, G. Su, M. Polacek, R. L. Kaufman, S. Tokumine, H. Hu, E. Buchatskaya, Y. Miao, M. Elhawaty, A. Siddhant, N. Tomasev, J. Xing, C. Greer, H. Miller, S. Ashraf, A. Roy, Z. Zhang, A. Ma, A. Filos, M. Besta, R. Blevins, T. Klimenko, C.-K. Yeh, S. Changpinyo, J. Mu, O. Chang, M. Pajarskas, C. Muir, V. Cohen, C. L. Lan, K. Haridasan, A. Marathe, S. Hansen, S. Douglas, R. Samuel, M. Wang, S. Austin, C. Lan, J. Jiang, J. Chiu, J. A. Lorenzo, L. L. Sjösund, S. Cevey, Z. Gleicher, T. Avrahami, A. Boral, H. Srinivasan, V. Selo, R. May, K. Aisopos, L. Hussenot, L. B. Soares, K. Baumli, M. B. Chang, A. Recasens, B. Caine, A. Pritzel, F. Pavetic, F. Pardo, A. Gergely, J. Frye, V. Ramasesh, D. Horgan, K. Badola, N. Kassner, S. Roy, E. Dyer, V. C. Campos, A. Tomala, Y. Tang, D. E. Badawy, E. White, B. Mustafa, O. Lang, A. Jindal, S. Vikram, Z. Gong, S. Caelles, R. Hemsley, G. Thornton, F. Feng, W. Stokowiec, C. Zheng, P. Thacker, Çağlar Ünlü, Z. Zhang, M. Saleh, J. Svensson, M. Bileschi, P. Patil, A. Anand, R. Ring, K. Tsihlas, A. Vezer, M. Selvi, T. Shevlane, M. Rodriguez, T. Kwiatkowski, S. Daruki, K. Rong, A. Dafoe, N. FitzGerald, K. Gu-Lemberg, M. Khan, L. A. Hendricks, M. Pellat, V. Feinberg, J. Cobon-Kerr, T. Sainath, M. Rauh, S. H. Hashemi, R. Ives, Y. Hasson, E. Noland, Y. Cao, N. Byrd, L. Hou, Q. Wang, T. Sottiaux, M. Paganini, J.-B. Lespiau, A. Moufarek, S. Hassan, K. Shivakumar, J. van Amersfoort, A. Mandhane, P. Joshi, A. Goyal, M. Tung, A. Brock, H. Sheahan, V. Misra, C. Li, N. Rakićević, M. Dehghani, F. Liu, S. Mittal, J. Oh, S. Noury, E. Sezener, F. Huot, M. Lamm, N. D. Cao, C. Chen, S. Mudgal, R. Stella, K. Brooks, G. Vasudevan, C. Liu, M. Chain, N. Melinkeri, A. Cohen, V. Wang, K. Seymore, S. Zubkov, R. Goel, S. Yue, S. Krishnakumaran, B. Albert, N. Hurley, M. Sano, A. Mohananey, J. Joughin, E. Filonov, T. Kępa, Y. Eldawy, J. Lim, R. Rishi, S. Badiezadegan, T. Bos, J. Chang, S. Jain, S. G. S. Padmanabhan, S. Puttagunta, K. Krishna, L. Baker, N. Kalb, V. Bedapudi, A. Kurzrok, S. Lei, A. Yu, O. Litvin, X. Zhou, Z. Wu, S. Sobell, A. Siciliano, A. Papir, R. Neale, J. Bragagnolo, T. Toor, T. Chen, V. Anklin, F. Wang, R. Feng, M. Gholami, K. Ling, L. Liu, J. Walter, H. Moghaddam, A. Kishore, J. Adamek, T. Mercado, J. Mallinson, S. Wandekar, S. Cagle, E. Ofek, G. Garrido, C. Lombriser, M. Mukha, B. Sun, H. R. Mohammad, J. Matak, Y. Qian, V. Peswani, P. Janus, Q. Yuan, L. Schelin, O. David, A. Garg, Y. He, O. Duzhyi, A. Älgmyr, T. Lottaz, Q. Li, V. Yadav, L. Xu, A. Chinien, R. Shivanna, A. Chuklin, J. Li, C. Spadine, T. Wolfe, K. Mohamed, S. Das, Z. Dai, K. He, D. von Dincklage, S. Upadhyay, A. Maurya, L. Chi, S. Krause, K. Salama, P. G. Rabinovitch, P. K. R. M, A. Selvan, M. Dektiarev, G. Ghiasi, E. Guven, H. Gupta, B. Liu, D. Sharma, I. H. Shtacher, S. Paul, O. Akerlund, F.-X. Aubet, T. Huang, C. Zhu, E. Zhu, E. Teixeira, M. Fritze, F. Bertolini, L.-E. Marinescu, M. Bölle, D. Paulus, K. Gupta, T. Latkar, M. Chang, J. Sanders, R. Wilson, X. Wu, Y.-X. Tan, L. N. Thiet, T. Doshi, S. Lall, S. Mishra, W. Chen, T. Luong, S. Benjamin, J. Lee, E. Andrejczuk, D. Rabiej, V. Ranjan, K. Styrc, P. Yin, J. Simon, M. R. Harriott, M. Bansal, A. Robsky, G. Bacon, D. Greene, D. Mirylenka, C. Zhou,

O. Sarvana, A. Goyal, S. Andermatt, P. Siegler, B. Horn, A. Israel, F. Pongetti, C.-W. L. Chen, M. Selvatici, P. Silva, K. Wang, J. Tolins, K. Guu, R. Yogev, X. Cai, A. Agostini, M. Shah, H. Nguyen, N. Donnaile, S. Pereira, L. Friso, A. Stambler, A. Kurzrok, C. Kuang, Y. Romanikhin, M. Geller, Z. Yan, K. Jang, C.-C. Lee, W. Fica, E. Malmi, Q. Tan, D. Banica, D. Balle, R. Pham, Y. Huang, D. Avram, H. Shi, J. Singh, C. Hidey, N. Ahuja, P. Saxena, D. Dooley, S. P. Potharaju, E. O'Neill, A. Gokulchandran, R. Foley, K. Zhao, M. Dusenberry, Y. Liu, P. Mehta, R. Kotikalapudi, C. Safranek-Shrader, A. Goodman, J. Kessinger, E. Globen, P. Kolhar, C. Gorgolewski, A. Ibrahim, Y. Song, A. Eichenbaum, T. Brovelli, S. Potluri, P. Lahoti, C. Baetu, A. Ghorbani, C. Chen, A. Crawford, S. Pal, M. Sridhar, P. Gurita, A. Mujika, I. Petrovski, P.-L. Cedoz, C. Li, S. Chen, N. D. Santo, S. Goyal, J. Punjabi, K. Kappaganthu, C. Kwak, P. LV, S. Velury, H. Choudhury, J. Hall, P. Shah, R. Figueira, M. Thomas, M. Lu, T. Zhou, C. Kumar, T. Jurdi, S. Chikkerur, Y. Ma, A. Yu, S. Kwak, V. Ähdel, S. Rajayogam, T. Choma, F. Liu, A. Barua, C. Ji, J. H. Park, V. Hellendoorn, A. Bailey, T. Bilal, H. Zhou, M. Khatir, C. Sutton, W. Rzadkowski, F. Macintosh, R. Vij, K. Shagin, P. Medina, C. Liang, J. Zhou, P. Shah, Y. Bi, A. Dankovics, S. Banga, S. Lehmann, M. Bredesen, Z. Lin, J. E. Hoffmann, J. Lai, R. Chung, K. Yang, N. Balani, A. Bražinskas, A. Sozanschi, M. Hayes, H. F. Alcalde, P. Makarov, W. Chen, A. Stella, L. Snijders, M. Mandl, A. Kärrman, P. Nowak, X. Wu, A. Dyck, K. Vaidyanathan, R. R, J. Mallet, M. Rudominer, E. Johnston, S. Mittal, A. Udathu, J. Christensen, V. Verma, Z. Irving, A. Santucci, G. Elsayed, E. Davoodi, M. Georgiev, I. Tenney, N. Hua, G. Cideron, E. Leurent, M. Alnahlawi, I. Georgescu, N. Wei, I. Zheng, D. Scandinaro, H. Jiang, J. Snoek, M. Sundararajan, X. Wang, Z. Ontiveros, I. Karo, J. Cole, V. Rajashekhar, L. Tumeh, E. Ben-David, R. Jain, J. Uesato, R. Datta, O. Bunyan, S. Wu, J. Zhang, P. Stanczyk, Y. Zhang, D. Steiner, S. Naskar, M. Azzam, M. Johnson, A. Paszke, C.-C. Chiu, J. S. Elias, A. Mohiuddin, F. Muhammad, J. Miao, A. Lee, N. Vieillard, J. Park, J. Zhang, J. Stanway, D. Garmon, A. Karmarkar, Z. Dong, J. Lee, A. Kumar, L. Zhou, J. Evens, W. Isaac, G. Irving, E. Loper, M. Fink, I. Arkatkar, N. Chen, I. Shafran, I. Petrychenko, Z. Chen, J. Jia, A. Levskaya, Z. Zhu, P. Grabowski, Y. Mao, A. Magni, K. Yao, J. Snaider, N. Casagrande, E. Palmer, P. Suganthan, A. Castaño, I. Giannoumis, W. Kim, M. Rybiński, A. Sreevatsa, J. Prendki, D. Soergel, A. Goedeckemeyer, W. Gierke, M. Jafari, M. Gaba, J. Wiesner, D. G. Wright, Y. Wei, H. Vashisht, Y. Kulizhskaya, J. Hoover, M. Le, L. Li, C. Iwuanyanwu, L. Liu, K. Ramirez, A. Khorlin, A. Cui, T. LIN, M. Wu, R. Aguilar, K. Pallo, A. Chakladar, G. Perng, E. A. Abellan, M. Zhang, I. Dasgupta, N. Kushman, I. Penchev, A. Repina, X. Wu, T. van der Weide, P. Ponnapalli, C. Kaplan, J. Simsa, S. Li, O. Dousse, F. Yang, J. Piper, N. Ie, R. Pasumarthi, N. Lintz, A. Vijayakumar, D. Andor, P. Valenzuela, M. Lui, C. Paduraru, D. Peng, K. Lee, S. Zhang, S. Greene, D. D. Nguyen, P. Kurylowicz, C. Hardin, L. Dixon, L. Janzer, K. Choo, Z. Feng, B. Zhang, A. Singhal, D. Du, D. McKinnon, N. Antropova, T. Bolukbasi, O. Keller, D. Reid, D. Finchelstein, M. A. Raad, R. Crocker, P. Hawkins, R. Dadashi, C. Gaffney, K. Franko, A. Bulanova, R. Leblond, S. Chung, H. Askham, L. C. Cobo, K. Xu, F. Fischer, J. Xu, C. Sorokin, C. Alberti, C.-C. Lin, C. Evans, A. Dimitriev, H. Forbes, D. Banarse, Z. Tung, M. Omernick, C. Bishop, R. Sterneck, R. Jain, J. Xia, E. Amid, F. Piccinno, X. Wang, P. Banzal, D. J. Mankowitz, A. Polozov, V. Krakovna, S. Brown, M. Bateni, D. Duan, V. Firoiu, M. Thotakuri, T. Natan, M. Geist, S. tan Girgin, H. Li, J. Ye, O. Roval, R. Tojo, M. Kwong, J. Lee-Thorp, C. Yew, D. Sinopalnikov, S. Ramos, J. Mellor, A. Sharma, K. Wu, D. Miller, N. Sonnerat, D. Vnukov, R. Greig, J. Beattie, E. Caveness, L. Bai, J. Eisenschlos, A. Korchemniy, T. Tsai, M. Jasarevic, W. Kong, P. Dao, Z. Zheng, F. Liu, F. Yang, R. Zhu, T. H. Teh, J. Sanmiya, E. Gladchenko, N. Trdin, D. Toyama, E. Rosen, S. Tavakkol, L. Xue, C. Elkind, O. Woodman, J. Carpenter, G. Papamakarios, R. Kemp, S. Kafle, T. Grunina, R. Sinha, A. Talbert, D. Wu, D. Owusu-Afriyie, C. Du, C. Thornton, J. Pont-Tuset, P. Narayana, J. Li, S. Fatehi, J. Wieting, O. Ajmeri, B. Uria, Y. Ko, L. Knight, A. Héliou, N. Niu, S. Gu, C. Pang,

Y. Li, N. Levine, A. Stolovich, R. Santamaria-Fernandez, S. Goenka, W. Yustalim, R. Strudel, A. Elqursh, C. Deck, H. Lee, Z. Li, K. Levin, R. Hoffmann, D. Holtmann-Rice, O. Bachem, S. Arora, C. Koh, S. H. Yeganeh, S. Põder, M. Tariq, Y. Sun, L. Ionita, M. Seyedhosseini, P. Tafti, Z. Liu, A. Gulati, J. Liu, X. Ye, B. Chrzaszcz, L. Wang, N. Sethi, T. Li, B. Brown, S. Singh, W. Fan, A. Parisi, J. Stanton, V. Koverkathu, C. A. Choquette-Choo, Y. Li, T. Lu, A. Ittycheriah, P. Shroff, M. Varadarajan, S. Bahargam, R. Willoughby, D. Gaddy, G. Desjardins, M. Cornero, B. Robenek, B. Mittal, B. Albrecht, A. Shenoy, F. Moiseev, H. Jacobsson, A. Ghaffarkhah, M. Rivière, A. Walton, C. Crepy, A. Parrish, Z. Zhou, C. Farabet, C. Radebaugh, P. Srinivasan, C. van der Salm, A. Fidjeland, S. Scellato, E. Latorre-Chimoto, H. Klimczak-Plucińska, D. Bridson, D. de Cesare, T. Hudson, P. Mendolicchio, L. Walker, A. Morris, M. Mauger, A. Guseynov, A. Reid, S. Odoom, L. Loher, V. Cotruta, M. Yenugula, D. Grewe, A. Petrushkina, T. Duerig, A. Sanchez, S. Yadlowsky, A. Shen, A. Globerson, L. Webb, S. Dua, D. Li, S. Bhupatiraju, D. Hurt, H. Qureshi, A. Agarwal, T. Shani, M. Eyal, A. Khare, S. R. Belle, L. Wang, C. Tekur, M. S. Kale, J. Wei, R. Sang, B. Saeta, T. Liechty, Y. Sun, Y. Zhao, S. Lee, P. Nayak, D. Fritz, M. R. Vuyyuru, J. Aslanides, N. Vyas, M. Wicke, X. Ma, E. Eltyshev, N. Martin, H. Cate, J. Manyika, K. Amiri, Y. Kim, X. Xiong, K. Kang, F. Luisier, N. Tripuraneni, D. Madras, M. Guo, A. Waters, O. Wang, J. Ainslie, J. Baldridge, H. Zhang, G. Pruthi, J. Bauer, F. Yang, R. Mansour, J. Gelman, Y. Xu, G. Polovets, J. Liu, H. Cai, W. Chen, X. Sheng, E. Xue, S. Ozair, C. Angermueller, X. Li, A. Sinha, W. Wang, J. Wiesinger, E. Koukoumidis, Y. Tian, A. Iyer, M. Gurumurthy, M. Goldenson, P. Shah, M. Blake, H. Yu, A. Urbanowicz, J. Palomaki, C. Fernando, K. Durden, H. Mehta, N. Momchev, E. Rahimtoroghi, M. Georgaki, A. Raul, S. Ruder, M. Redshaw, J. Lee, D. Zhou, K. Jalan, D. Li, B. Hechtman, P. Schuh, M. Nasr, K. Milan, V. Mikulik, J. Franco, T. Green, N. Nguyen, J. Kelley, A. Mahendru, A. Hu, J. Howland, B. Vargas, J. Hui, K. Bansal, V. Rao, R. Ghiya, E. Wang, K. Ye, J. M. Sarr, M. M. Preston, M. Elish, S. Li, A. Kaku, J. Gupta, I. Pasupat, D.-C. Juan, M. Someswar, T. M., X. Chen, A. Amini, A. Fabrikant, E. Chu, X. Dong, A. Muthal, S. Buthpitiya, S. Jauhari, N. Hua, U. Khandelwal, A. Hitron, J. Ren, L. Rinaldi, S. Drath, A. Dabush, N.-J. Jiang, H. Godhia, U. Sachs, A. Chen, Y. Fan, H. Taitelbaum, H. Noga, Z. Dai, J. Wang, C. Liang, J. Hamer, C.-S. Ferng, C. Elkind, A. Atias, P. Lee, V. Listík, M. Carlen, J. van de Kerkhof, M. Pikus, K. Zaher, P. Müller, S. Zykova, R. Stefanec, V. Gatsko, C. Hirnschall, A. Sethi, X. F. Xu, C. Ahuja, B. Tsai, A. Stefanoiu, B. Feng, K. Dhandhania, M. Katyal, A. Gupta, A. Parulekar, D. Pitta, J. Zhao, V. Bhatia, Y. Bhavnani, O. Alhadlaq, X. Li, P. Danenberg, D. Tu, A. Pine, V. Filippova, A. Ghosh, B. Limonchik, B. Urala, C. K. Lanka, D. Clive, Y. Sun, E. Li, H. Wu, K. Hongtongsak, I. Li, K. Thakkar, K. Omarov, K. Majmundar, M. Alverson, M. Kucharski, M. Patel, M. Jain, M. Zabelin, P. Pelagatti, R. Kohli, S. Kumar, J. Kim, S. Sankar, V. Shah, L. Ramachandruni, X. Zeng, B. Bariach, L. Weidinger, T. Vu, A. Andreev, A. He, K. Hui, S. Kashem, A. Subramanya, S. Hsiao, D. Hassabis, K. Kavukcuoglu, A. Sadovsky, Q. Le, T. Strohman, Y. Wu, S. Petrov, J. Dean, and O. Vinyals. Gemini: A family of highly capable multimodal models, 2025. URL https://arxiv.org/abs/2312.11805.

S. C. Thakur. Foundation models for time series forecasting. *International IT Journal of Research, ISSN: 3007-6706*, 2(4):144–156, 2024.

R. Thapa, B. He, M. R. Kjaer, H. M. Iv, G. Ganjoo, E. Mignot, and J. Zou. Sleepfm: Multi-modal representation learning for sleep across brain activity, ecg and respiratory signals. In *International Conference on Machine Learning*, pages 48019–48037. PMLR, 2024.

Y. Tian, D. Krishnan, and P. Isola. Contrastive representation distillation. *arXiv*, 2019. doi: 10.48550/arxiv.1910.10699. URL https://arxiv.org/abs/1910.10699.

A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

N. S. V. S. Venkata. Ai model deployment in healthcare: Mlops innovations and challenges. *Journal ID*, 9471:1297, 2025.

V. Vijayan, J. P. Connolly, J. Condell, N. McKelvey, and P. Gardiner. Review of wearable devices and data collection considerations for connected health. *Sensors*, 21(16):5589, 2021.

R. S. Vulcan, S. André, and M. Bruyneel. Photoplethysmography in normal and pathological sleep. *Sensors*, 21(9):2928, 2021.

M. Wornow, Y. Xu, R. Thapa, B. Patel, E. Steinberg, S. Fleming, M. A. Pfeffer, J. Fries, and N. H. Shah. The shaky foundations of large language models and foundation models for electronic health records. *npj digital medicine*, 6(1):135, 2023.

F. Xu, H. Uszkoreit, Y. Du, W. Fan, D. Zhao, and J. Zhu. Explainable ai: A brief survey on history, research areas, approaches and challenges. In *Natural language processing and Chinese computing: 8th cCF international conference, NLPCC 2019, dunhuang, China, October 9–14, 2019, proceedings, part II 8*, pages 563–574. Springer, 2019.

M. A. Xu, J. Narain, G. Darnell, H. Hallgrimsson, H. Jeong, D. Forde, R. Fineman, K. J. Raghuram, J. M. Rehg, and S. Ren. Relcon: Relative contrastive learning for a motion foundation model for wearable data, 2025a. URL https://arxiv.org/abs/2411.18822.

M. A. Xu, G. Narayanswamy, K. Ayush, D. Spathis, S. Liao, S. A. Tailor, A. Metwally, A. A. Heydari, Y. Zhang, J. Garrison, et al. Lsm-2: Learning from incomplete wearable sensor data. *arXiv preprint arXiv:2506.05321*, 2025b.

H. Yuan, S. Chan, A. P. Creagh, C. Tong, A. Acquah, D. A. Clifton, and A. Doherty. Self-supervised learning for human activity recognition using 700,000 person-days of wearable data. *NPJ digital medicine*, 7(1):91, 2024a.

H. Yuan, T. Plekhanova, R. Walmsley, A. C. Reynolds, K. J. Maddison, M. Bucan, P. Gehrman, A. Rowlands, D. W. Ray, D. Bennett, J. McVeigh, L. Straker, P. Eastwood, S. D. Kyle, and A. Doherty. Self-supervised learning of accelerometer data provides new insights for sleep and its association with mortality. *npj Digital Medicine*, 7(1):86, may 2024b. doi: 10.1038/s41746-024-01065-0. URL http://dx.doi.org/10.1038/s41746-024-01065-0.

Y. Zhang, K. Ayush, S. Qiao, A. A. Heydari, G. Narayanswamy, M. A. Xu, A. A. Metwally, S. Xu, J. Garrison, X. Xu, et al. Sensorlm: Learning the language of wearable sensors. *arXiv preprint arXiv:2506.09108*, 2025.

W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong, et al. A survey of large language models. *arXiv preprint arXiv:2303.18223*, 1(2), 2023.

X. Zhu, J. Li, Y. Liu, C. Ma, and W. Wang. A survey on model compression for large language models. *Transactions of the Association for Computational Linguistics*, 12:1556–1577, 2024.