

Building a Foundation Model for Human Health

Executive Summary

A **foundation model for human health** is an ambitious project to collect comprehensive, longitudinal health data from thousands of individuals and train AI systems that can model human physiology, behavior, and disease. Recent advances in wearables, multi-omics, and AI suggest that such a “digital health twin” is within reach. This report reviews cutting-edge efforts (2023–2024) toward this goal, highlighting large-scale sensor studies, multi-modal cohorts, partnership opportunities, technical infrastructure, and key application areas.

In the past year, companies like Google and Apple have developed **large sensor foundation models** using data from hundreds of thousands of people. Google’s **Large Sensor Model (LSM)** was trained on 40 million hours of smartwatch data from 165,000 users, demonstrating that performance scales predictably with data and compute ¹. Apple researchers built a generalizable accelerometer-based model using 20 million minutes of wearable data from 172,000 participants ². These self-supervised models capture rich health signals and can be adapted for many downstream tasks.

The foundation model vision requires massive, diverse datasets. We survey major **cohort initiatives**: the US **All of Us** program (targeting 1 million multi-ethnic participants with genomics, electronic health records (EHR), surveys, and wearable data), the UK **Biobank** (500,000 participants with genomic and phenotypic data, including a subset with 7-day accelerometry ³), and the new UK **Our Future Health** cohort (over 1 million already enrolled toward 5 million total, making it the largest longitudinal study globally ⁴). We also highlight focused sensor trials like Samsung’s **HEARTBEAT** study (10,000 people wearing Galaxy Watch 6 for a year to monitor cardiovascular health ⁵). Partnering with **device manufacturers** (Apple, Fitbit/Google, Oura, Withings, etc.) provides access to validated wearable hardware and user bases, accelerating data collection. Emerging smart devices – from connected **toothbrushes** monitoring oral hygiene to **smart toilets** analyzing stool and urine – open new avenues for passive health monitoring ⁶.

A critical challenge is acquiring **high-quality, high-frequency biomarker data** within a feasible budget (~\$10,000 per person-year). Innovative strategies include performing lab tests and multi-omics assays in **low-cost regions**. For example, advanced diagnostic labs in India can run comprehensive test panels at a small fraction of US costs – one case documented an entire battery of 72 tests for \$50 in India versus an estimated \$10,000 in the US ⁷. By leveraging such cost efficiencies and economies of scale, a \$100M budget could cover 10,000 participants with dense monthly blood tests, whole-genome sequencing, microbiome profiling, and continuous monitoring devices. Careful **budget modeling** is needed to allocate funds across devices (wearables, sensors), assays (lab tests, sequencing), personnel, and compute resources, optimizing data yield per dollar.

Ensuring the **public benefit** of a health foundation model is paramount. We discuss open-science practices such as data sharing under controlled access and open-source model release. Large cohorts like UK Biobank already operate as open-access resources for bona fide researchers ⁸. For a new initiative, adopting participant-centric governance and privacy protection is key. Data must be rigorously de-identified

and secured. Techniques like **federated learning** can allow models to train on distributed data without raw data leaving secure sites, safeguarding privacy while still leveraging “big data” ⁹. All software and AI models should use permissive licenses to maximize public utility, and research findings should be published openly. This approach aligns with the **open-source ethos**, enabling global researchers to build on the work while upholding participants’ privacy and trust.

On the **technical front**, we outline the AI and infrastructure required to build a reasoning-capable digital health twin. A multi-modal model would ingest continuous streams from wearables (heart rate, activity, sleep, etc.), periodic clinical test results, medication records, lifestyle logs, and more. Self-supervised learning (e.g. masked signal modeling) has been shown to produce powerful representations from raw sensor data ¹⁰. Such a model can be **pre-trained** to impute missing data, detect anomalies, and forecast health trajectories. We propose integrating a **deep reinforcement learning (RL)** agent on top of this predictive model to enable decision-making and “what-if” scenario testing. Reinforcement learning can leverage patient-specific simulations to find optimal interventions – akin to how doctors adjust treatments over time – by trialing actions in silico and learning policies that improve long-term outcomes ¹¹. This could empower the digital twin to not only passively predict health outcomes but also actively recommend personalized lifestyle changes or therapies. Achieving this requires significant computational power. NVIDIA’s cutting-edge **GPU infrastructure** (such as the DGX systems powering the Cambridge-1 supercomputer with 400+ petaflops for AI ¹²) is well-suited to train large bio-medical models. NVIDIA’s platforms like **Omniverse** further enable creating realistic digital simulations – for example, combining medical imaging data with virtual environments to test how an intervention might affect a patient’s organ systems ¹³.

Use Case: Semaglutide and Multidimensional Health Outcomes. To illustrate the model’s potential, we examine data on **semaglutide (Ozempic)**, a GLP-1 agonist drug, and how a health twin could model its effects. Clinical studies show semaglutide profoundly alters several health domains. It induces weight loss primarily by **reducing appetite and caloric intake**, leading to a ~15% body weight reduction in 68 weeks ¹⁴ ¹⁵. Participants on semaglutide have significantly lower inflammation markers (e.g. C-reactive protein) and improved metabolic profiles compared to placebo ¹⁶. These changes translate into hard outcomes: the SELECT trial (2023) demonstrated a **20% reduction in major cardiovascular events** (heart attack, stroke, CV death) in overweight patients treated with semaglutide versus placebo ¹⁷. There is also emerging evidence of cognitive benefits – a real-world study in diabetic patients found those on semaglutide had a significantly lower hazard of developing Alzheimer’s disease compared to other treatments ¹⁸. Semaglutide’s multimodal impact (appetite, weight, inflammation, vascular health, and possibly neuroprotection ¹⁹) exemplifies the complexity that a foundation model must capture. A digital twin could integrate an individual’s data to predict how starting semaglutide might affect their future weight trajectory, inflammation levels, **cancer risk**, and organ health. Notably, early research suggests GLP-1 drugs may confer ~40% risk reduction in obesity-related cancers beyond what weight loss alone would predict ²⁰, potentially via direct anti-inflammatory and metabolic effects. By learning from thousands of treated and untreated individuals, the model could discern these drug effects and even simulate long-term outcomes (e.g. 10-year cardiovascular or cognitive risk) for a given patient profile. This would enable clinicians and patients to make more informed decisions about therapies like semaglutide in the context of overall health.

In conclusion, building a foundation model for human health is a grand but achievable endeavor that synergizes large-scale cohort data, wearable and smart device technology, advanced AI modeling, and open collaborative science. Early projects by leading tech firms demonstrate the feasibility of training generalist

health models from raw sensor data. By expanding data collection to rich multi-omics and continuous monitoring – and doing so in a cost-efficient, inclusive manner – we can create a truly representative dataset of human health. With robust HPC infrastructure and state-of-the-art algorithms (self-supervised learning, federated learning, and reinforcement learning), this data can yield a **digital health twin** that not only mirrors an individual's current health status but can reason about future outcomes and optimal interventions. Such a model holds promise to transform preventive medicine, allowing us to detect disease shifts early, personalize treatments (e.g. choosing who benefits most from a drug like semaglutide), and ultimately improve health outcomes globally. The next sections delve into each aspect of this vision, citing recent evidence and practical considerations for turning it into reality.

Introduction

Modern healthcare is on the cusp of a transformation driven by data and artificial intelligence. In other domains, **foundation models** – large AI models trained on diverse data – have achieved remarkable general capabilities (e.g. GPT-4 in language). The concept of a foundation model for human health is to similarly learn a broad, holistic representation of human physiology and disease from extensive multimodal data ¹. The model would function as a “digital twin,” a computational mirror of an individual's health state that can be used for risk prediction, early detection of issues, and testing interventions *in silico*. Building such a model requires integrating data across **sensors, clinical records, and omics**, at a scale and granularity not seen in traditional medical studies. This report explores the recent progress toward this goal and the key components needed – from large cohort datasets and device partnerships to AI techniques and computing infrastructure – to make a health foundation model a reality.

Recent Advances in Health Foundation Models (2023–2024)

In the last two years, researchers have begun creating general-purpose models for wearable and health data, signaling the emergence of health foundation models. Notably, **Google Health** and **Apple** have leveraged their vast device ecosystems in pioneering studies:

- *Google's Large Sensor Model (LSM)*: Google researchers published a 2024 study on scaling self-supervised models for consumer health data ²¹ ¹. They assembled what is described as “*the largest wearable dataset published to date*” – over **40 million hours** of continuous smartwatch sensor data from 165,000 participants wearing Fitbit Sense 2 or Pixel Watch devices ²² ²³. By training a transformer-based model on this data (with a masked time-series modeling objective), they created a **Large Sensor Model** capable of imputing missing sensor data and forecasting future signals. Crucially, they demonstrated classic scaling law behavior: as they increased the training data and model size, performance on downstream tasks improved, with up to 38% gains over conventional approaches ²². This suggests that, analogous to text and image models, **more data and larger models yield better understanding in the health domain**. Google's LSM was evaluated on tasks like activity recognition and showed strong generalization, indicating it learned meaningful physiological representations from the raw streams.
- *Apple's Foundation Models for Wearables*: Apple's 2025 report (work done in 2024) described training a general **accelerometer-based health model** ²⁴. Using **20 million minutes** of Apple Watch accelerometer data from ~172,000 participants in the Apple Heart and Movement Study, they applied knowledge distillation from a parallel photoplethysmography (PPG) model to create a high-

fidelity accelerometer encoder ². The resulting model could accurately infer heart rate and variability from motion data alone, and predicted a wide array of health targets, confirming it as a “*generalist foundation model*” ². This is significant because accelerometers are ubiquitous (low-power and present in virtually all wearables), so a powerful accelerometer-based model can extend advanced health analytics to even simple devices. Apple’s work highlights how combining modalities during training (in their case, motion plus PPG) produces richer representations – a principle that can extend to combining wearable data with other data types.

Beyond Big Tech, academic efforts are also contributing. For example, open-source initiatives in 2023 introduced foundation models for human activity recognition that developers can build on ²⁵. Similarly, researchers have proposed motion foundation models for mental health and other applications ²⁶. These efforts reinforce a trend: **health data is moving toward foundation models** that capture generic patterns (heart rhythms, activity profiles, sleep cycles, etc.) which can be fine-tuned for many specific predictive tasks with minimal labeled data ²¹ ²⁷.

In summary, the past year’s developments by Google, Apple, and others have shown that large-scale self-supervised learning on wearable and health datasets is not only feasible but yields powerful general models. These serve as early building blocks for a comprehensive health twin. The next sections discuss how to gather the requisite data, ensure diversity, and integrate other crucial modalities (clinical, biochemical, behavioral) to broaden these models’ scope.

Large-Scale Cohorts and Data Sources for Model Training

Any foundation model is only as good as the data it’s trained on. To comprehensively model human health, we need **large, diverse, longitudinal datasets** that capture different ages, ethnicities, and health states. Fortunately, several national and global cohort initiatives are actively building such datasets:

- **All of Us Research Program (USA):** This NIH-led project aims to enroll **1 million Americans** with deliberate oversampling of underrepresented groups. Enrollees contribute EHR data, survey answers (on lifestyle, diet, etc.), physical measurements, and biospecimens for genomic and other analyses. Many also opt-in to wearable device data sharing (the program provides Fitbit integration) ²⁸. Already, All of Us has released initial data on hundreds of thousands of participants. For example, one analysis included **5.9 million person-days of Fitbit data** (steps and heart rate) from 6,042 participants collected over a median of 4 years ²⁹. This shows the scale of longitudinal behavior data available. As All of Us grows, it will provide an unparalleled multi-modal dataset linking continuous lifestyle data with medical outcomes in a diverse population.
- **UK Biobank (UK):** UK Biobank is a well-established cohort of **500,000** participants recruited in 2006–2010, now with ~15 years of follow-up. It includes extensive baseline phenotyping, genotype data, and linkage to health records. Crucially, UK Biobank has pioneered incorporating new data modalities: a subset of ~100,000 participants wore a wrist accelerometer for 1 week, generating an open activity dataset ³. Another 100,000 underwent advanced imaging (MRI of brain, heart, etc.). UK Biobank’s open-access model (data is available to qualified researchers worldwide) ⁸ and rich dataset make it a cornerstone for population health modeling. Any foundation health model would benefit from pre-training or validating on this cohort. For instance, algorithms can be checked against Biobank’s known outcomes (e.g. incident cardiovascular events, cancers) to ensure the model’s risk predictions are calibrated.

- **Our Future Health (UK):** This is a newer UK initiative scaling beyond Biobank. Launched in 2022, **Our Future Health** plans to recruit up to **5 million** adults. Remarkably, it has already enrolled over **1 million participants** by late 2024 ⁴, making it the largest prospective cohort of its kind. Participants provide blood samples (for biomarker and genetic analysis) and health data, with repeat assessments over time. The study focuses on earlier detection and prevention of diseases like dementia, cancer, and heart disease ³⁰. The massive size and breadth of this cohort will allow training AI models with unprecedented statistical power. It also presents an opportunity to work with device companies – e.g. a subset of the 5 million could be equipped with wearables – to create the largest sensor dataset integrated with clinical data.
- **International Cohort Collaborations:** Beyond the US/UK, many countries have large biobanks or cohort studies (e.g. the 500K-participant **China Kadoorie Biobank**, 150K **Japan Biobank**, etc.). The **International HundredK+ Cohorts Consortium (IHCC)** links many of these to foster global analyses. A foundation model project could leverage data federation across multiple cohorts to improve diversity. For instance, one could train initial models on UK/US data and further validate or fine-tune on Asian or African cohorts, ensuring the model generalizes across ancestries and environments.
- **Disease-Specific and High-Intensity Cohorts:** Some studies focus on specific groups but collect extremely dense data. For example, the MIMIC critical care database (although hospital-based) or studies like the Framingham Heart Study (with multigenerational cardiovascular data) offer valuable focused insights. Additionally, Verily (Alphabet’s life sciences arm) conducted the **Project Baseline** study with about 2,500 participants, each extensively characterized (wearables, quarterly clinic visits, imaging, genomics) over four years ³¹ ³². Similarly, a 108-person pilot by Snyder et al. (Stanford) collected multi-omics monthly for wellness monitoring. While smaller, these datasets are deep and can help validate model predictions against gold-standard clinical measures.

A strategy for building the foundation model is to **integrate these data sources**. This could mean literally pooling data (where permissible) or using transfer learning between them. For instance, one might pre-train on a broad population dataset like Our Future Health, then fine-tune on high-quality multi-omic data from a subcohort to imbue the model with biochemical insight. It will be important to align data schemas and ontologies (for clinical variables, etc.) across cohorts – an effort that initiatives like OHDSI and FHIR standards can assist. Overall, today’s cohorts provide an excellent starting point: on the order of 10 million individuals globally have some form of research-grade longitudinal data that could feed into a health foundation model.

Integrating Wearables and Smart Devices: Partnerships and Data Streams

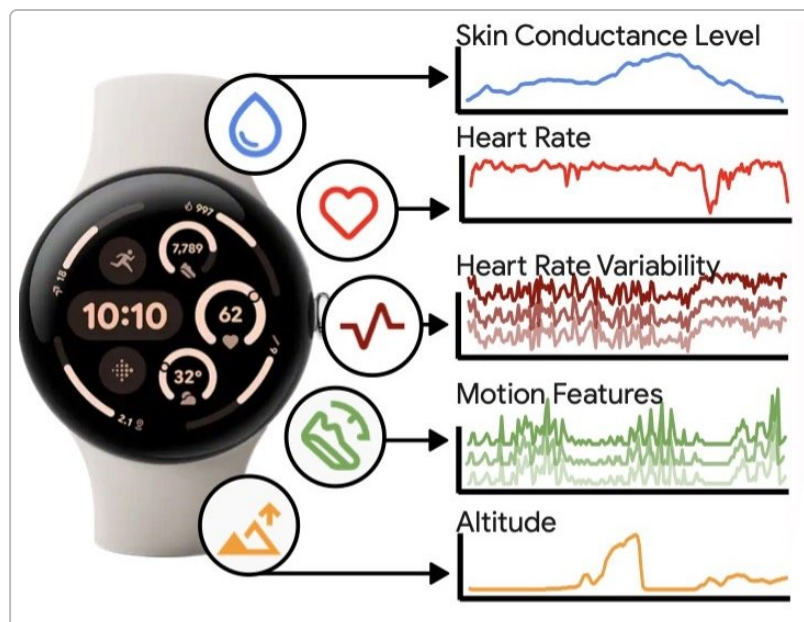
A distinguishing feature of the envisioned project is the use of **continuous wearable sensors and “internet of things” health devices** to capture lifestyle and physiologic data with high resolution. This section discusses the types of devices, partnerships to obtain them, and examples of their use in large studies.

Wearable Devices for Vital Signs and Activity: Modern smartwatches and fitness bands pack an array of sensors – photoplethysmography (for heart rate and blood oxygen), accelerometers and gyroscopes

(movement and orientation), temperature sensors, even electrodermal activity (stress/sweat response) in some models. These devices can record 24/7 data. Partnering with major manufacturers can ensure access to reliable devices and even existing user data (with consent). For example, **Apple** could contribute Apple Watch data via its ongoing Apple Heart & Movement Study, which already enrolled hundreds of thousands of users through the Apple Research app. **Google (Fitbit)** has a huge user base and is developing personalized health insights for Fitbit users ³³ ³⁴. **Samsung** has engaged in research partnerships (e.g. the HEARTBEAT trial providing Galaxy Watches to 10,000 participants ⁵). **Oura** (maker of a smart ring) and **Whoop** (fitness band) are startups known for advanced sleep and recovery metrics, and they have collaborated with academic researchers on studies (such as using Oura rings for illness detection). **Withings** provides smart scales, blood pressure monitors, and even smart thermometers – these could feed data on weight, body composition, blood pressure, and fever patterns.

Partnering with these companies can take multiple forms: bulk purchasing devices at discounted research rates; direct collaboration where the company helps recruit users (e.g., sending study invites to their app users); or data-sharing agreements to use de-identified historical data. All approaches should be pursued to maximize data volume. It's worth noting that many people already own such devices – a study could allow participants to use their own device (BYOD approach) to reduce cost, by building an app or platform for data donation.

Data from wearables is high-dimensional and rich. For instance, a smartwatch can yield heart rate readings every few minutes, continuous activity counts, and sleep stage estimates each night. When multiplied by 10,000+ individuals over a year, this becomes a **big data challenge** but also a treasure trove for modeling. Self-supervised learning can uncover latent patterns: circadian rhythm signatures, fitness levels, early signs of illness (e.g. elevated resting heart rate), etc. These signals often precede or supplement traditional clinical metrics. For example, subtle changes in sleep or resting heart rate variability may indicate emerging issues like infection or overtraining.



Examples of health signals captured by a modern smartwatch (Google Pixel Watch): continuous sensor streams include skin conductance (stress response), heart rate and its variability, motion features, and even altitude

changes. Foundation models can learn from such multi-modal time-series to derive meaningful health indicators

23 10 .

Beyond Wrist Wearables – Smart Devices in Daily Life: To truly quantify health behaviors, one should also capture data from everyday activities. **Connected toothbrushes** are an interesting example – they monitor brushing frequency, duration, and sometimes technique. Research trials have shown that smart toothbrushes can improve oral health by giving feedback ³⁵. In a health cohort, a connected toothbrush could provide data on oral hygiene habits, which relate to overall health (poor oral health is linked to cardiovascular disease and dementia risk). **Smart scales** (e.g. Withings Body+) can transmit daily weight and body composition. **Continuous glucose monitors (CGM)**, while typically for diabetics, are increasingly used in wellness contexts; they can log glucose fluctuations in response to meals, giving insight into metabolic health. **Blood pressure cuffs** can be used at home (some are Bluetooth-enabled) to track hypertension management. Even **smartphones** themselves, via apps, can collect useful data: geolocation (for activity space and environmental exposure), app usage patterns (potentially related to mental health), and questionnaires or cognitive tests administered regularly.

Two particularly novel devices deserve mention:

- **Smart Toilets:** A research team at Stanford has prototyped a “precision health toilet” that automatically analyzes urine and stool for biomarkers ^{6 36}. The unit has cameras and chemical sensors; it can measure things like stool form (for GI health), detect blood or protein in urine, and even identify users by an “anal print” for data attribution (famously earning an Ig Nobel prize). In a pilot, the smart toilet successfully captured meaningful health data from volunteers ^{37 38}. While still experimental, this concept could become a game-changer for continuous monitoring of renal function, hydration, gut health, etc., without active user intervention. Scaling up would require manufacturing robust versions of these toilets or retrofittable kits. If achievable, participants could get real-time health feedback from their bathrooms, and the study would receive frequent biochemical data (potentially daily). Privacy and acceptance are concerns (some may be uncomfortable with a camera in the toilet), but if positioned as a medical device with proper safeguards, many would see the value for health tracking.

- **Ambient Home Sensors:** These include devices like air quality monitors (tracking pollution or allergens in the participant’s environment), smart speakers or other IoT devices that could, with permission, track health-relevant behaviors (e.g. cough frequency detection via smart speaker, or fridge sensors tracking diet). **Bed sensors** (like the Google Nest Hub’s sleep sensing or pressure mats) can unobtrusively record sleep duration and quality. **Smartphone-based cognitive tests** can be thought of as sensors for cognitive health (for example, a weekly memory game app to detect subtle cognitive changes).

Incorporating a broad range of devices yields a **multi-sensor fusion problem** – but also an opportunity for the foundation model to learn cross-correlations (e.g., how does late-night smartphone use or poor sleep reflect in next day’s blood sugar or mood?). Each device partner can contribute a layer to the overall picture of health.

Partnership Opportunities: The report has already mentioned potential partners (Apple, Google, Samsung, Oura, Withings, etc.). Engaging them could involve co-designing the study protocol to ensure device data quality and user compliance. Many startups are also emerging with unique health tech – for example, companies making **wearable patches for continuous hormone or nutrient monitoring**, or **smart contact lenses** measuring intraocular pressure or glucose. Academic engineering departments could collaborate by providing prototypes (e.g., a university might have a new portable EEG headband or a

pollution sensor that could be deployed). The project should remain agile in adopting new devices as they become available, ensuring the data stays on the cutting edge of what can be measured outside a lab.

To manage the device data, one needs a robust **digital platform**. All participant devices would ideally connect to a unified app or hub (on the participant's phone or home WiFi), which uploads data to the study cloud in real time. Platforms like **Open mHealth** have defined standard schemas for wearable data streams, which should be utilized for interoperability. The study app can also handle participant engagement: reminding people to wear their devices, performing device troubleshooting, and even rewarding compliance. A high adherence to wearing devices and using smart tools is critical – the best sensor is useless if left on the shelf. Pilot testing and human-centered design can refine the approach to ensure that the technology seamlessly blends into participants' lives.

In summary, harnessing **wearable and smart device data** is a linchpin of the foundation health model effort. By partnering with hardware makers and leveraging participants' own devices, the project can gather continuous, real-world health data at an unprecedented scale and frequency. This provides the raw material for the AI models to learn the signatures of health and disease in everyday life.

High-Frequency Lab Testing and Multi-Omics Data

While wearables capture physiology and behavior, a comprehensive health model also needs **internal biological data** – the kind typically obtained via laboratory tests or molecular profiling. This includes blood chemistry, genomics, proteomics, metabolomics, microbiome analyses, and more. Collecting such data regularly (e.g., monthly) from 10,000 participants for a year is a formidable undertaking, but it is key to linking the external sensor data with internal health states (e.g., linking a rise in night-time heart rate to an increase in inflammatory markers or linking activity patterns to metabolic health).

Core Clinical Lab Tests: Each participant could undergo routine blood tests at high frequency (monthly or quarterly). These might include complete blood count, metabolic panel, lipid panel, HbA1c, CRP (inflammation marker), hormone levels (thyroid, cortisol), vitamin levels, and so on. Over a year, this generates a time-series for each marker, capturing short-term fluctuations and longer-term trends. For example, one might observe seasonal variation in vitamin D, or rising fasting glucose as someone progresses toward diabetes. Laboratory infrastructure is needed to process ~10,000 samples per month. Partnering with large diagnostic companies (Quest, LabCorp in the US, or SRL in Asia) or setting up a centralized lab facility is required. Here, **location matters**: performing these tests in countries where costs are lower can stretch the budget. As noted earlier, **India, Malaysia, Thailand** and others have world-class diagnostic labs with much lower prices. A striking anecdote showed an American getting 72 tests for \$50 in India, vs tens of thousands of dollars in the U.S. ⁷. The gap is huge. By either recruiting participants in those regions or shipping samples there (the latter could be complicated by logistics and regulations), the project can drastically reduce per-test costs. In India, large chains like Dr. Lal PathLabs or Thyrocare have optimized, automated facilities. For example, a thyroid panel or liver function test may cost only a few dollars locally. One approach could be to set up regional lab hubs – e.g., one in South Asia, one in Southeast Asia – where samples from participants (even if participants live elsewhere) are sent for processing. However, transporting biological samples internationally in bulk adds complexity (customs, stability). Alternatively, one could recruit a substantial number of participants from those regions outright, achieving both diversity and cost efficiency, as labor and overhead costs for the entire study could be lower there.

Genomics: Whole genome sequencing (WGS) for 10,000 people is desirable to capture genetic risk factors and enable polygenic risk scores. Costs of WGS have dropped markedly; high-coverage genomes can be done for a few hundred dollars each in bulk. Companies like Illumina and BGI, as well as genome centers in Singapore or India, could perform this. There are also smaller “exotic” genomic assays: e.g., epigenetic clocks (to measure biological aging via methylation patterns) or telomere length assays. These could be done at baseline and end-of-study to see how intensive monitoring and interventions impact biological aging.

Multi-Omics Profiling: The \$10,000 per person budget suggests room for extensive molecular profiling. Studies like the Arivale wellness cohort showed the feasibility of measuring thousands of molecules per person ³⁹. In Arivale’s case, participants had **genomics, longitudinal metabolomics, proteomics, clinical labs, gut microbiome sequencing, wearable data, and lifestyle surveys** ³⁹. We propose a similar multi-omic approach: - **Metabolomics:** Using mass spectrometry to measure hundreds of metabolites in blood (and possibly urine). This provides a biochemical snapshot of metabolism, diet, gut microbial byproducts, etc. For example, it can detect elevated blood fatty acids or markers of oxidative stress. - **Proteomics:** Profiling circulating proteins (e.g., using aptamer or antibody arrays) can capture the state of pathways (inflammation, hormonal signals, cardiac markers). There are technologies like Olink or Somalogic that can measure 1000–7000 proteins from a small sample. Doing this quarterly could reveal dynamic changes in, say, inflammatory cytokines or cardiac troponin levels, even within normal ranges, that correlate with sensor data or predict upcoming illness. - **Microbiome:** Periodic stool samples for 16S rRNA or metagenomic sequencing will characterize the gut microbiota, which influences and reflects health (implicating digestion, immunity, even mood). If a smart toilet is implemented, it could potentially grab samples automatically; if not, participants could mail stool samples using provided kits. - **Other ‘omics:** Depending on budget, transcriptomics (RNA-seq of blood for gene expression), epigenomics (methylation patterns), and metabolomic profiles of other fluids (urine, saliva) could be added. These deeper omics yield high-dimensional data that, while costly and complex, could allow the model to link gene expression changes or microbiome shifts to phenotypic changes (like weight gain or high stress periods).

Ensuring Quality and Consistency: High-throughput omics assays must be done in batches and with calibration. The study should allocate part of the budget to robust quality control, including replicate samples and reference controls. This ensures that observed changes are biological, not technical artifacts. Another aspect is **data harmonization** – since different assays produce very different data types (counts, concentrations, sequences), the model’s architecture will need to accommodate these. Likely, specialized sub-models or embeddings can be created for each data type (like a “metabolomic encoder”), which then feed into the broader foundation model.

Lab Infrastructure in Low-Cost Regions: As explicitly requested, focusing on places like **India, Malaysia, Thailand** for lab work can drastically reduce costs while maintaining quality. India, for instance, has CAP-accredited labs (e.g., MedGenome for genomics ⁴⁰) and has even become a medical tourism hub for affordable diagnostics. By outsourcing or establishing a lab in these countries, the study could perhaps do *more* tests per person for the same money. For example, instead of quarterly metabolomics in the US, monthly metabolomics in India might be feasible. This requires collaboration with local healthcare providers and possibly navigating local regulations for sample collection if participants are not local. However, given the global nature of many cohort studies now, it’s an approach worth pursuing. Some projects have collected samples in the US and shipped to BGI in China or similar for cheap whole-genome sequencing – models for such collaboration exist.

Participant Experience and Logistics: To collect frequent samples, one must make it convenient. Options include mobile phlebotomy (a phlebotomist comes to the participant's home or workplace for blood draws) or local lab visits (partner with local clinics or lab chains in participants' cities). Mobile phlebotomy services exist in the US and could be contracted. Alternatively, in a region like India, one could integrate with the existing network of labs where patients commonly schedule home sample collection via an app (a practice already popular there). Proper scheduling, reminders, and follow-up for missed samples are needed to minimize data gaps. Less invasive methods can also be employed for some metrics: finger-prick microsampling (dried blood spots) can be self-collected by participants and mailed for certain analyses, though volume is limited. Urine samples could be collected in mailed cups for lab analysis of things like cortisol or metabolite excretion if a smart toilet is not used.

In conclusion, robust **lab and omics data** will complement the wearable data by providing ground-truth measures of internal health states. The fusion of these – for instance, correlating a person's activity patterns with their blood sugar and metabolomic profile – is what will enable the foundation model to truly understand health-disease transitions. It is challenging but feasible within the budget if we smartly leverage global resources and new technologies. The next section addresses how to optimize these extensive data collections within budgetary constraints.

Budget Optimization for High-Quality Data Acquisition

Executing a 10,000-person intensive study for one year with a budget of \$10k each (total ~\$100 million) demands careful financial planning. Here we outline a **budget model** and ways to optimize costs while maintaining data quality:

1. Devices and Equipment: Equipping participants with wearables and sensors is a upfront cost. However, many may already own a compatible device. We can budget, say, \$500 per person for devices (this could cover a smartwatch, an Oura ring, possibly a BP cuff or scale). Bulk purchasing from manufacturers could reduce this – e.g., negotiating with Apple or Fitbit for a research discount. If participants use their own devices, we might instead give an incentive or data stipend. We should also budget for replacement devices (some will be lost or broken over the year). Estimated allocation: ~\$3–4 million for devices and replacements (assuming some cost offset by existing device use).

2. Lab Tests and Omics: This will likely be the largest expense category. If we assume monthly blood chemistry panels (~\$50 each in bulk) and quarterly expanded panels (vitamins, hormones etc.), plus baseline and final more extensive testing, per person might be ~\$500 for standard clinical labs in low-cost settings. Whole genome sequencing might be \$300–\$500 each by 2025 rates. High-dimensional assays like metabolomics/proteomics are still pricey: maybe \$200 per assay per timepoint. If done quarterly, that's \$800 per person per assay type; doing both metabolomics and proteomics would be \$1,600. Microbiome sequencing might be \$100 per sample; done twice, \$200. Summing: roughly \$500 (clinical labs) + \$400 (genome amortized over year) + \$1,600 (multi-omics) + \$200 (microbiome) = \$2,700 per person. This could be lowered with bulk deals or if not all participants get all assays (one could subset for omics if needed). But given the budget, \$2–3k for assays is reasonable and leaves room to potentially do more omics (or cover analysis costs).

3. Personnel and Operations: A project of this scale needs a team: project managers, study coordinators, data engineers, lab technicians, and an ethics/compliance team. We might allocate ~\$10M (10% of budget)

for staffing over the project's duration. Using partners and outsourcing can reduce some staff load (e.g., contracting a phlebotomy service rather than hiring phlebotomists directly).

4. Participant Recruitment and Retention: It's crucial to recruit 10,000 people and keep them engaged. Marketing and outreach costs (advertising the study, community engagement) might require a few million. We should also budget participant **incentives**. For a demanding study, compensation might be around \$500–\$1000 per participant for the year (in installments or device gifts etc.). Let's say \$750 average – that's \$7.5M. If altruism and the cool factor of the project attract some for less, we could save, but planning for incentive costs is wise to ensure representation from all socio-economic levels.

5. Data Systems and Cloud Infrastructure: Storing and processing continuous data, genomic data, etc., will incur substantial cloud computing costs (and later AI training costs). We might set aside ~\$5M for cloud storage and data platform development (with security). Some tech partners might sponsor cloud credits (for example, if using Google Cloud or AWS, they might support a high-profile health project).

6. Analysis and AI Development: The budget should also account for the R&D: salaries for data scientists, computational costs for model training (GPUs time). Training a large model on this data might take a cluster of GPUs for weeks – that could be on the order of hundreds of thousands of dollars. Given we plan to use NVIDIA infrastructure possibly via existing supercomputers or partnerships, we may get this subsidized. Still, a line item of a few million for analysis ensures we can fully exploit the data. We note that using open-source tools and academic collaborations can offset some analysis costs.

7. Contingency: Always plan ~10% contingency for unexpected costs (regulatory delays, needing to replace a vendor, etc.).

To **optimize costs**: - Leverage **existing networks**: e.g., if All of Us or Our Future Health already have participants and data, coordinate rather than duplicate efforts. Possibly piggyback additional sensor deployment on those cohorts. - **Volume contracts**: Commit to high volumes with labs (as mentioned, use labs in Asia for cheap rates) ⁷. For devices, get sponsorship – e.g., device companies might donate or subsidize devices in exchange for co-authorship or data insights. - **Phased approach**: Perhaps start with a pilot of 1,000 participants to refine protocols, avoiding expensive mistakes on the full 10k. Then scale up. This ensures the large investment is not wasted on a flawed design. - **Automation**: Use tech to reduce manual labor – e.g., automated digital questionnaires instead of phone interviews, AI for data cleaning, etc. A lot of health study cost is in managing data; good engineering can reduce that. - **Open-source software**: Use open tools for data collection and analysis (no expensive proprietary software licenses). The project can adopt or build on existing open platforms for study management and thereby save development cost.

By carefully allocating resources as above, the project can remain within the ~\$100M budget while achieving its ambitious data collection goals. It's about balancing the expenditures between breadth (more participants) and depth (more data per participant). Given the target is a foundation model, we lean slightly toward **depth of data** on this fixed cohort, to maximize what the AI can learn, rather than recruiting far more people with sparse data. The chosen 10k size is a compromise to capture population heterogeneity yet allow intense monitoring.

Open-Source Practices and Data Privacy Considerations

Building a foundational health model in an open, ethical manner is as important as the technical achievements. We must ensure that the **data is used for public benefit**, participants' privacy is protected, and resulting models and insights are shared responsibly. Here, we outline practices and licensing strategies aligned with those goals:

Open Data (with Safeguards): Ideally, the collected dataset (or at least a significant portion of it) should be made available to the broader research community under appropriate agreements. Models like UK Biobank show this is feasible – researchers apply and agree to terms, then gain access to de-identified data for approved analyses ⁸. For our study, we can implement a data enclave or cloud-based workbench where approved researchers can run analyses without downloading sensitive data, adding an extra layer of security. At minimum, derived data and summary statistics (e.g., genome-wide association summary stats, trained model embeddings) should be openly published. This ensures **maximal scientific yield** from the data beyond what the core study team can achieve. However, true open release of raw data to the public is not advisable given privacy issues; a controlled access model is better.

Licensing of Models and Code: Any AI models, algorithms, or software developed should be released under an open-source license (such as Apache 2.0 or MIT). This allows others to use and build upon the models in their own applications, maximizing the public good. There has been discussion of “responsible AI licenses” (like the RAIL license) for health or sensitive domains – but a straightforward open license combined with an ethical use policy might suffice. The **foundation model** itself (the digital health twin) once trained could be shared as: (a) a set of pretrained weights for research (with sensitive components perhaps withheld if there's risk of re-identification, although model weights generally don't leak raw data); and/or (b) as an interactive service or API for individuals and clinicians to query their own data against. It's important that this doesn't become a proprietary tool locked behind paywalls – since the project is presumably publicly funded or philanthropically funded, the outputs should remain a public resource.

Participant Privacy and Consent: All participants must give informed consent that clearly explains data uses, sharing, and how privacy will be preserved. Given the intrusive data being collected (genetic data, constant monitoring, etc.), we should implement state-of-the-art privacy measures. Data will be stored with coded IDs, and direct identifiers (name, address, etc.) kept separate or hashed. When releasing data, **de-identification** procedures like removing or generalizing dates and locations will be applied. However, with granular longitudinal data, complete anonymity is hard – hence the controlled access model. We might also explore **differential privacy** techniques for releasing statistical results (ensuring that no individual can be re-identified from aggregated model outputs). Another approach is **federated learning**, as mentioned earlier: rather than centralizing all data, one could train models at the data's location and only share model parameters. In practice for our study, true federation might not be necessary because participants consent to data pooling, but federated or distributed training could further reduce privacy risks and also engage other data holders (like hospital networks who could keep their patient data on-site while contributing to model training).

A recent review on federated foundation models in healthcare emphasizes that this combination can “harness [model] analytical power while safeguarding the privacy of sensitive medical data” ⁴¹ ⁹. We can take a hybrid approach: centralize the richly consented study data, but also incorporate federated learning from external datasets that can't be shared (like EHRs from a health system). This would enrich the model without breaching those systems' privacy regulations.

Ethical Oversight: An ethics board or governance committee including participant representatives should oversee data use. This body can evaluate proposals for data access and ensure they align with participant interests. Transparency with participants is key – e.g., provide newsletters on how the data is being used, what findings emerge, etc. This maintains trust, which is crucial if we seek to extend the study or recruit future cohorts.

Open-Source Community and Collaboration: We can foster an ecosystem around the project by releasing intermediate data and challenges. For example, an open challenge on predicting certain health events from the wearable data could engage external teams (similar to how PhysioNet challenges work). This crowdsourced innovation can accelerate progress and also test the generalizability of the model approaches.

Licenses Ensuring Public Benefit: There are examples in software of licenses that prevent misuse – for example, a license that forbids using the software to harm individuals or for surveillance. We might consider a clause that the foundation model cannot be used to deny insurance or employment, etc. However, enforcing such clauses is tricky. It may be better to advocate for policies (laws/regulations) that guard against misuse of health AI, rather than rely purely on license. In any case, the data and tools will be intended for public health benefit, and this should be baked into agreements with any commercial partners too (e.g., device partners shouldn't get exclusive rights to the data).

In summary, by adopting an **open-source, privacy-by-design approach**, we ensure the project's outcomes are widely accessible and ethically managed. The success of a foundation health model should be measured not only in accuracy but in how it empowers researchers and individuals around the world to improve health. As we proceed to building the AI, keeping these principles in focus will help maintain public trust and maximize impact.

AI Model Architecture and NVIDIA Infrastructure for a Digital Health Twin

With data in hand, the next step is developing the **AI architecture** that will learn from it. The goal is a “reasoning-capable” model – one that not only predicts correlations but can infer cause-effect to some extent and simulate scenarios like a true **digital twin**. We break down the components needed and how modern AI hardware (especially NVIDIA's) will be leveraged.

Multi-Modal Model Architecture: The model must handle different data types: time-series signals from wearables, static information (genetics, demographics), event data (medical diagnoses or interventions), and possibly image data (if we include imaging studies). A promising approach is a **modular neural network** where each modality has a dedicated encoder that transforms raw input into embeddings, and then a fusion network integrates these. For example: - A **time-series transformer** (similar to models used for wearables ¹⁰) that takes in streams like heart rate, activity, sleep and produces a latent representation of the person's physiologic state over time. - A **graph neural network or embedding for clinical events** from EHR (representing diagnoses, medications, etc. in a temporal sequence). - A **genomic embedding** (e.g., using a polygenic risk scoring or a smaller network that encodes key genetic risk factors). - **Omics embeddings** for high-dimensional data like metabolomics – perhaps using an autoencoder trained on that data to reduce dimensionality.

These can be concatenated or integrated via cross-attention mechanisms into a unified latent space. The model can be designed as a **Transformer** across time that can attend to different modalities' signals at each time point. Some signals are continuous (wearables) and some are sparse (lab tests monthly); the model's timeline would have "observations" of various types and use a masking approach to deal with missing values.

Google's work on the Large Sensor Model is instructive – they used a masking and imputation pretext task ¹⁰. We can extend that: during pre-training, randomly mask portions of data (like some vital sign readings or some lab results) and train the model to predict them from the context. This forces the model to learn the relationships (e.g., if someone's activity increases and diet changes, how does that reflect in their lab tests? The model might learn to reconstruct a missing CRP value from recent sleep and heart rate patterns, etc.). Such **self-supervised objectives** are key for a foundation model, as they make use of the plentiful unlabeled data.

Generative and Discriminative Capacities: Once trained, the model should be able to do **generative tasks** (simulate forward in time, impute missing data) as well as **discriminative tasks** (classify risk of disease, etc.). For example, given a person's data up to today, generate the probable trajectories of health metrics for the next month. This is essentially a sequence prediction and is useful for anticipating problems. The discriminative side might be a classifier head that the model feeds into to predict, say, the likelihood of developing diabetes in the next year. Because the foundation model has a rich picture of the person, these predictions can be more accurate than traditional risk models.

Reinforcement Learning for Reasoning: One limitation of pure predictive models is that they operate in a passive mode. To have a reasoning-capable twin, we incorporate **Deep Reinforcement Learning (DRL)**. In practice, this means creating an environment where the model can try out different actions (interventions) on a simulated patient (the model of that patient) and learn policies that optimize certain outcomes. For instance, actions could be: recommend a diet change, start a medication, suggest more sleep, etc., and the reward could be a healthier outcome (like reduced predicted 10-year cardiovascular risk). The "state" is the latent health state of the digital twin, which evolves based on both naturally simulated progression and applied actions. By training an agent (using algorithms like DQN or policy gradient methods) in this environment, we get a **policy module** that can suggest interventions.

A simpler application of RL in our context is sequential decision support: for chronic conditions that require continuous management (like titrating insulin for diabetes or antihypertensives for blood pressure), an RL agent could learn to adjust doses based on the data trends. Indeed, previous work in healthcare has used RL to propose treatment policies in ICU or for sepsis management ¹¹. A review in 2024 highlighted that "*RL leverages patient-data to create personalized treatment plans, improving outcomes*" ¹¹. Our digital twin could extend that concept beyond single diseases: it could be an RL-based coach that monitors all aspects and gives holistic advice (e.g., it might "decide" that improving sleep and reducing stress will yield better overall reward than solely focusing on one biomarker).

NVIDIA Infrastructure: Training such a complex model with potentially billions of parameters and multi-modal data is computationally intensive. We anticipate using **NVIDIA GPU clusters** for this. NVIDIA has been actively supporting healthcare AI – for example, the **Cambridge-1 supercomputer** (built by NVIDIA in the UK) is dedicated to healthcare research and can deliver 400+ petaflops for AI ¹². Our project could partner to utilize Cambridge-1 or similar HPC resources. Additionally, the **DGX systems** (NVIDIA's AI servers) could be deployed exclusively for this project's use if budget allows, or cloud GPU services can be rented.

Efficient training frameworks (Mixed precision, etc.) will be used to handle the large throughput of time-series data.

NVIDIA's software stacks like **Clara** might be useful for certain aspects (Clara has frameworks for genomics and medical imaging). Also, **NVIDIA Omniverse** was mentioned – Omniverse could allow visualization and simulation, for instance, one could create a virtual human model and simulate how interventions play out, though that's more in the realm of physics-based simulation. However, a company (Medical IP) has already connected medical imaging data to Omniverse for creating digital twin visuals ¹³. For us, Omniverse might be a way to combine the data-driven AI twin with a 3D visualization of anatomy or to integrate physics-based models of, say, blood flow or biomechanics with our AI model's predictions.

We also plan to incorporate **NVIDIA's AI libraries** (like NeMo for large language models, if we integrate any text or conversational interface to the twin, or CUDA optimizations for custom ops in the model). By leaning on NVIDIA's expertise and possibly their engineers via a partnership, we can ensure the model is optimized to run efficiently.

Model Evaluation and Reasoning: A reasoning-capable model should be able to explain or justify its suggestions to gain trust. We might integrate methods for **explainable AI**, such as attention weight interpretation (e.g., highlighting which data points the model relied on for a given prediction) or using a companion large language model that can translate model findings into human-readable advice. Google's mention of building a personal health LLM for Fitbit users ³³ hints that combining sensor data with LLMs to communicate insights is on the horizon. Our foundation model could similarly feed summary features into a generative LLM that interacts with patients or doctors in natural language ("Your digital twin noticed that whenever your weekly exercise drops below 60 minutes, your blood sugar the next week rises by 10%. It suggests keeping activity above that threshold to maintain glycemic control."). This marries deep analytics with accessible reasoning.

To summarize, the AI architecture will be a **multi-modal deep learning model with self-supervised pre-training**, augmented by a **reinforcement learning agent** for decision recommendations. Training it will rely on **massive parallel computing** on GPUs, exemplified by NVIDIA's systems. The result will be an AI system that can simulate an individual's health – essentially a **digital health twin** that can predict and also prescribe. In the next section, we examine in more detail how such a model would apply to analyzing complex interventions like drug effects, tying together many of the concepts discussed so far.

Application Example: Modeling Drug Effects (Semaglutide and Beyond)

One of the powerful uses of a health foundation model is to assess and predict the effects of interventions, such as medications, on a whole-person level. Traditional clinical trials focus on specific outcomes and may miss broader impacts. A comprehensive model can consider multi-system effects and individual variability. We illustrate this with **semaglutide (Ozempic)**, a GLP-1 agonist originally for diabetes, now widely used for weight loss. Semaglutide's effects span multiple domains – metabolic, inflammatory, cardiovascular, possibly cognitive – making it an ideal test case for a digital twin's reasoning.

Weight Loss and Behavior: Semaglutide leads to substantial weight loss by strongly suppressing appetite. In a pivotal trial, participants on semaglutide 2.4 mg weekly lost ~14.9% of body weight in 68 weeks on

average, versus ~2.4% for placebo ⁴². The drug “stems weight loss from a reduction in energy intake owing to decreased appetite” ¹⁵. Our model, given data from many individuals on semaglutide, would capture how their dietary intake, step counts, and satiety signals changed. For instance, wearable data might show reduced eating frequency or shorter meal durations (if using a smart fork or logging diet in an app) and increased daily activity as weight decreases. The model could identify responders vs. non-responders: perhaps some people’s neural reward signals (inferred indirectly via things like heart rate variability during meals or sleep quality) adapt differently. It might also notice behavioral compensations (some might eat fewer calories but more high-fat foods, etc.). By learning these patterns, the twin could predict how a given person’s behavior and weight would respond to semaglutide before they even start – enabling personalized decisions (maybe someone with certain behavioral or microbiome profile will only lose 5%, not worth it, whereas another will lose 20%).

Inflammation and Metabolic Health: Weight loss typically reduces systemic inflammation. Indeed, the semaglutide trial noted a greater drop in C-reactive protein (CRP) levels in the treated group ¹⁶. Our multi-omics data would likely show semaglutide’s effect on inflammatory cytokines, liver enzymes (since fatty liver improves with weight loss), and insulin sensitivity markers. In fact, semaglutide improves glycemic control even in non-diabetics to some extent. The model would link the time-course of weight loss with improvements in labs: for example, predicting that after 3 months on the drug, a person’s CRP might drop from 5 mg/L to 2 mg/L, HDL cholesterol might increase, etc. If a person had baseline high inflammation, the model might project an extra benefit in inflammation reduction (since obesity-related inflammation is being removed). On the other hand, if someone has a different source of inflammation (like autoimmune disease), the model might predict less change, isolating the drug effect versus other factors.

Cardiovascular Outcomes: The **SELECT trial** (published 2023) provided strong evidence that semaglutide *reduces cardiovascular events by ~20%* in overweight people without diabetes ¹⁷. This is a composite of heart attacks, strokes, etc. A digital twin trained on data including who had events could potentially identify *why* this reduction happened. It might see that semaglutide users have lower blood pressure (some weight loss effect), improved cardiac function due to lower fat mass, and less progression of atherosclerosis markers. By having continuous data (e.g., maybe some participants wear blood pressure monitors or ECG patches), the model could detect subtle improvements in cardiovascular physiology (like heart rate recovery after exercise improving, or fewer overnight heart rate spikes indicating lower sympathetic drive). These intermediate phenotypes can strengthen the causal picture that weight loss via semaglutide leads to a healthier cardiovascular system. The model could simulate: “what if this patient stays on semaglutide for 5 years – how much will their 10-year heart attack risk drop?” using not just population average (20%) but their personal data (if they have a strong family history, maybe their absolute risk reduction is still meaningful but not as high, etc.). This helps in shared decision-making – maybe someone with moderate obesity but very high heart risk stands to gain a lot, whereas someone with obesity but already great metabolic profile might gain more in quality of life than in hard outcomes.

Cognitive and Alzheimer’s Disease: Fascinatingly, research is emerging that GLP-1 agonists may benefit brain health. Preclinical studies in mice show that semaglutide reduces amyloid plaques and neuroinflammation, improving cognitive function ⁴³. And a human retrospective study found **~40–67% lower hazard of Alzheimer’s diagnosis** in diabetic patients on semaglutide vs other meds ¹⁸. Two large trials (EVOKE and EVOKE+ for early Alzheimer’s) are underway to test semaglutide in that population ¹⁹⁴⁴. Our model, combining cognitive testing, MRI (if available for a subset), and biomarkers, could learn whether semaglutide has measurable cognitive effects in our generally healthy cohort. For instance, maybe participants on semaglutide have slightly better sleep and hence better cognitive performance on weekly

phone games. Or their inflammatory markers (like TNF-alpha) drop, which is linked to slowed cognitive decline. By integrating all this data, the digital twin might predict long-term brain outcomes: e.g., that 15 years on semaglutide (for weight maintenance) could translate into a 30% reduction in Alzheimer's risk for those with obesity and insulin resistance. If someone has APOE4 genotype (higher AD risk), the model might particularly encourage interventions like semaglutide that improve metabolic health, as mounting evidence connects metabolic and cognitive trajectories.

Cancer Risk: Obesity is a risk factor for several cancers (colon, breast, endometrial, etc.). Weight loss should reduce that risk. Early studies, as cited, indicate GLP-1 drugs might **reduce obesity-related cancer incidence by ~41%** beyond just the weight loss effect ²⁰. Perhaps GLP-1s have direct effects on cell growth pathways or reduce inflammation which contributes to carcinogenesis ⁴⁵. Our foundation model would track any cancer diagnoses (though in one year, few would occur – but maybe over longer term or via linkage to records we'd see signals). More immediately, it could track precancerous changes, like improvement in fatty liver (reducing liver cancer risk) or reduction in insulin (hyperinsulinemia can promote cancer). With multi-omics, the model might detect changes in certain metabolic signatures associated with cancer risk (for example, lower levels of certain amino acids that in metabolomic studies correlate with higher cancer risk). Thus the model could quantify how much a given individual's cancer-risk biomarkers improve with semaglutide. Eventually, with enough follow-up, it could predict actual cancer incidence differences. This is of high interest because there have been concerns as well (rodent studies of GLP-1 agonists showed thyroid C-cell tumors) ⁴⁶, but human data so far are reassuring ⁴⁷. Our model could contribute to this safety monitoring by looking for any early signals of increased tumor biomarkers (none are known, but perhaps something like calcitonin levels for thyroid, etc., could be checked).

Polypharmacy and Interactions: Many individuals take multiple drugs (e.g., blood pressure meds, antidepressants). The digital twin can factor those in. Maybe it learns, for example, that semaglutide combined with an SGLT2 inhibitor (another diabetes drug) has an additive effect on weight and heart health. Or that people on certain psychiatric meds (which often cause weight gain) lose a bit less weight on semaglutide than others. This granular knowledge is usually beyond clinical trials (which exclude many such patients), but a broad observational foundation model could learn it. It can then make personalized predictions: "Given you are on medication X that increases appetite, your expected weight loss on semaglutide is 10% instead of 15%. Still worthwhile, and it will help counteract X's side effects."

In conclusion, the **digital health twin** armed with multi-modal data can become a powerful tool to simulate drug effects for an individual. It moves us toward a future of in silico trials and N-of-1 prediction. We specifically discussed semaglutide due to its prominence as a multi-system drug; the same framework applies to other interventions (e.g., predicting benefits of a new cholesterol drug on not just cholesterol but on exercise capacity and inflammation, or predicting side effects like how an antihistamine might subtly affect sleep cycles captured on a wearable). By validating the model's drug effect predictions against known trial results (like SELECT for heart disease, etc.), we build confidence in its accuracy. Eventually, regulators might even consider such validated models to support drug approvals or recommendations by providing evidence in populations underrepresented in trials.

Conclusion

Building a foundation model for human health is an interdisciplinary Grand Challenge – blending cutting-edge technology, biomedical science, and population-scale data collection. In this report, we surveyed recent breakthroughs that make this endeavor plausible now: enormous wearable datasets enabling self-

supervised learning of health signals ²², large cohorts and biobanks providing diverse data ⁴, and advanced AI methods that can unify multi-modal information and even reason about interventions. We discussed how a concerted project could be organized with 10,000 participants deeply monitored for a year, leveraging partnerships with device manufacturers and low-cost lab infrastructure to gather an unprecedented dataset spanning from daily step counts to whole-genome sequences.

Critically, we emphasized principles of openness and ethics – the resulting models and data should become a public good, catalyzing research globally while safeguarding individual privacy. The model itself, envisioned as a digital health twin, would not just predict one outcome but serve as a personal health oracle of sorts: integrating data on heart rhythm irregularities, blood biomarkers, genomics, lifestyle and more to flag risks early and suggest tailored preventive actions.

Our detailed example with semaglutide illustrated the power of such an approach: a single medication's ripple effects across weight, inflammation, cardiovascular health, and cognition can be captured and personalized ¹⁵ ¹⁷ ¹⁸. A foundation model could tell a patient not only “this drug can help you lose weight” but also “it might improve your 5-year heart health and even has potential cognitive benefits, given your profile” – essentially providing a 360-degree view of health improvement.

The development of the foundation model will also push the frontiers of AI. Handling long-term longitudinal data and reasoning about interventions (with reinforcement learning agents) will yield new algorithms and possibly novel insights into human biology. For instance, the model might uncover previously unknown early-warning patterns for diseases (maybe a combination of subtle changes in resting heart rate and cytokine levels that precedes autoimmune flares, for example). These could become new digital biomarkers ⁶ guiding future care.

We stand at a convergence of trends: wearable adoption is global and growing, omics costs are falling, and AI can finally ingest and make sense of the deluge of health data. Large tech companies and research consortia alike recognize the opportunity – as evidenced by Google's health LLMs and open-source model initiatives, and by international projects like Our Future Health scaling up to millions. The next step – and the focus of this report – is to tie these threads together into an integrated effort to build a comprehensive health model.

The benefits of success are immense. For individuals, it could mean having a “virtual twin” doctor alongside their real providers, constantly learning from their data to give informed, personalized advice. For health systems, it shifts the paradigm to prevention and precision – catching things early (because the model predicts them) and targeting interventions to those who truly benefit. For science, it creates a platform to test hypotheses at scale in silico, greatly accelerating discovery. And for society, especially if done openly, it democratizes health insights – everyone could eventually use an app powered by such a model to improve their well-being, not just those who can afford extensive medical workups.

Of course, challenges remain. Ensuring privacy and avoiding misuse of such powerful data is paramount. We must also be cautious to validate the model's recommendations with clinical evidence and retain the human touch in healthcare – the model augments but doesn't replace medical professionals. Participants must be treated as partners in research, with transparency and respect.

In closing, the vision of a health foundation model aligns with the broader movement toward **precision medicine and preventive care**. It leverages the best of technology in service of human health. The recent

papers and projects we reviewed show that pieces of this puzzle are already being put in place ²² ² ⁵ . With a focused initiative – properly funded and ethically conducted – we can assemble those pieces into a functioning reality. The coming years could then see the emergence of digital health twins guiding each of us, helping to extend healthy lifespan and improve quality of life, fulfilling one of the ultimate promises of modern biomedical science.

Sources: ¹ ² ⁵ ⁴ ²⁹ ³ ⁷ ⁹ ¹¹ ¹⁵ ¹⁷ ¹⁸ ¹⁹ ²⁰ ⁶

¹ ¹⁰ ²¹ ²² ²³ ²⁷ **Scaling wearable foundation models**

<https://research.google/blog/scaling-wearable-foundation-models/>

² ²⁴ ²⁶ **Wearable Accelerometer Foundation Models for Health via Knowledge Distillation - Apple Machine Learning Research**

<https://machinelearning.apple.com/research/wearable-accelerometer-foundation-models>

³ ⁸ ²⁹ **aaimecine.org**

<https://aaimecine.org/wp-content/uploads/2024/12/ADoherty-UK-Biobank-Data-and-Wearables.pdf>

⁴ ³⁰ **Our Future Health becomes world's largest research programme of its kind | Science Media Centre**

<https://www.sciencemediacentre.org/our-future-health-becomes-worlds-largest-research-programme-of-its-kind/>

⁵ **Samsung, Huma and Tulane University's HEARTBEAT 10,000 Study Enrolls First Patient**

<https://www.huma.com/resources/samsung-huma-and-tulane-universitys-heartbeat-10-000-study-enrolls-first-patient>

⁶ ³⁶ ³⁷ ³⁸ **'Smart toilet' monitors for signs of disease**

<https://med.stanford.edu/news/all-news/2020/04/smart-toilet-monitors-for-signs-of-disease.html>

⁷ **Indian Healthcare System: 50 dollars in India vs 10,000 dollars in the US for medical tests: How India's health care on a budget is gaining global attention | - The Times of India**

<https://timesofindia.indiatimes.com/life-style/health-fitness/health-news/50-dollars-in-india-vs-10000-dollars-in-the-us-for-medical-tests-how-indias-health-care-on-a-budget-is-gaining-global-attention/articleshow/114553143.cms>

⁹ ⁴¹ **Open challenges and opportunities in federated foundation models towards biomedical healthcare | BioData Mining | Full Text**

<https://biodatamining.biomedcentral.com/articles/10.1186/s13040-024-00414-9>

¹¹ **A Primer on Reinforcement Learning in Medicine for Clinicians - PMC**

<https://pmc.ncbi.nlm.nih.gov/articles/PMC11599275/>

¹² **Nvidia finds a home for the UK's most powerful supercomputer | AI Business**

<https://aibusiness.com/verticals/nvidia-finds-a-home-for-the-uk-s-most-powerful-supercomputer>

¹³ **Medical IP Is Developing Medical Digital Twins for NVIDIA Omniverse**

<https://www.medicaldevicetechnology.com/view/medical-ip-digital-twins-nvidia-omniverse>

¹⁴ ¹⁵ ¹⁶ ⁴² **Semaglutide effective for weight loss in non-diabetic adults, research suggests - The Pharmaceutical Journal**

<https://pharmaceutical-journal.com/article/news/semaglutide-effective-for-weight-loss-in-non-diabetic-adults-research-suggests>

¹⁷ **News Details**

<https://www.novonordisk.com/news-and-media/news-and-ir-materials/news-details.html?id=166301>

- 18 Associations of semaglutide with first-time diagnosis of Alzheimer's disease in patients with type 2 diabetes: Target trial emulation using nationwide real-world data in the US - PubMed
<https://pubmed.ncbi.nlm.nih.gov/39445596/>
- 19 44 evoke and evoke+: design of two large-scale, double-blind, placebo-controlled, phase 3 studies evaluating efficacy, safety, and tolerability of semaglutide in early-stage symptomatic Alzheimer's disease | Alzheimer's Research & Therapy | Full Text
<https://alzres.biomedcentral.com/articles/10.1186/s13195-024-01666-7>
- 20 45 Ozempic-Like GLP-1 Drugs May Lower Cancer Risk Up to 41%
<https://www.healthline.com/health-news/glp-1-drugs-lower-obesity-related-cancer-risk>
- 25 Self-supervised learning for human activity recognition using ...
<https://www.nature.com/articles/s41746-024-01062-3>
- 28 Resources for Using Fitbit Data - User Support
<https://support.researchallofus.org/hc/en-us/articles/20281023493908-Resources-for-Using-Fitbit-Data>
- 31 32 Google's Verily Launches 10,000-Person Study to 'Create a Map of Human Health' - BioSpace
<https://www.biospace.com/google-s-verily-launches-10-000-person-study-to-create-a-map-of-human-health>
- 33 34 Google expanding its healthcare AI ambitions in 2024
<https://www.fiercehealthcare.com/ai-and-machine-learning/google-scales-generative-ai-healthcare-advances-assist-clinicians-and-give>
- 35 Artificial-Intelligence-Based Smart Toothbrushes for Oral Health and ...
<https://www.mdpi.com/2673-947X/5/1/5>
- 39 Multiomic signatures of body mass index identify heterogeneous health phenotypes and responses to a lifestyle intervention | Nature Medicine
https://www.nature.com/articles/s41591-023-02248-0?error=cookies_not_supported&code=456275d4-4635-4593-9953-8f39ee027d05
- 40 Know About Genome Sequencing Labs in India - MedGenome
<https://diagnostics.medgenome.com/about-us/>
- 43 Semaglutide improves cognitive function and neuroinflammation in APP/PS1 transgenic mice by activating AMPK and inhibiting TLR4/NF-κB pathway - PubMed
<https://pubmed.ncbi.nlm.nih.gov/40151913/>
- 46 The new weight-loss drugs and cancer
<https://www.fredhutch.org/en/news/center-news/2024/03/the-new-weight-loss-drugs-and-cancer.html>
- 47 Can Semaglutide (Ozempic) Cause Cancer? - GoodRx
<https://www.goodrx.com/ozempic/does-semaglutide-cause-cancer?srsId=AfmBOoo7fj8aqDJneoDIixzySC5emzCou2kJaL3dYlk3uNDfKXIppqBa>