# AI tool predicts over 1,000 diseases years before they happen — and more are on the way

**Computer models are showing the potential to draw upon a person's clinical records and current health data to forecast diseases and treatment outcomes. Could this presage a new era of preventive medicine?**

**By Michael Eisenstein**

Foundation models integrate multiple layers of human data to predict future health and disease.

CREDIT: SLEEPYFELLOW/ALAMY

Last September, a team led by researchers at the European Bioinformatics Institute and German Cancer Research Center made a media splash with a *Nature* paper describing an artificial intelligence (AI) model that can simulate individual health trajectories decades into the future on the basis of electronic health records to predict the likelihood that a given disease might strike. The Delphi-2M model was trained on 400,000 medical records from the UK Biobank, identifying patterns of onset and progression that could be used to forecast up to 1,200 diseases 20 years in advance in both British and Danish cohorts. This model draws on details from a person's medical history while also taking into account health modifiers such as smoking, body mass index and alcohol consumption.

Over the past five years, several pioneering groups have explored AI-based models trained on massive population datasets to set up systems that aim to forecast an individual's future health. Some initiatives are academic, like Delphi-2M, but startups are also exploring this space, such as Tel Aviv-based Pheno.AI or Bethesda-based RespondHealth, alongside major tech and pharma players like Google and Roche.

Such AI tools could transform medicine, enabling health systems and governments to better manage finite resources and companies to run more effective clinical trials. Other models aim to create a virtual representation of an individual's clinical history and current health status: a 'digital twin' that can be used to test potential therapies and guide treatment planning. "The thing that excites me the most is being able to personalize deeply to the patient, where it's no longer a decision that's being made in the abstract," says Rahul Shah, a software engineer at electronic health record company Epic Systems.

Clinical researchers now have mountains of rich biomedical data at their disposal in the form of electronic health records (EHRs). These are commonplace in many countries, providing a detailed timeline of a disease's initial diagnosis and progression by tracking every patient's medical encounters and their treatment outcomes. For example, Shah's team at Epic developed a resource called Cosmos that contains deidentified records from over 300 million patients. In parallel, government-backed research programs like the UK Biobank and the US National Institutes of Health-led All of Us study, which have each collected data from over 500,000 participants to date, have complemented EHRs with clinically relevant information such as genomic data and other physiological metrics.

But for all the digital wealth amassed in EHRs, these have yet to translate into actionable predictive insights — to the considerable frustration of precision medicine enthusiasts. "I feel like it's kind of borderline criminal that we spend all this time collecting data and at the moment do so very little with it," says Richard Dobson, a researcher at King's College London who co-developed a predictive model called Foresight, first described in a 2024 paper.

Modern deep learning tools have now given the medical informatics community the chance to mine EHRs and extract insights that can be used to forecast a patient's health or disease trajectory in the real world. Most of these tools are based on transformers, the foundational deep learning architecture that underlies widely used large language models (LLMs) like ChatGPT. Transformers break datasets down into fundamental units called 'tokens', which they analyze with guidance from an 'attention' mechanism that allows the algorithm to prioritize relevant elements over extraneous information. In ChatGPT, these tokens are words and phrases, but for a medical forecasting model, these tokens could also be genome sequences, clinical tests, diagnostic codes or diverse other clinical data points. Such transformer-based analysis of tokens makes it possible to find hidden patterns in the data — for example, recurring links between particular risk factors or diagnostic measurements and future clinical outcomes.

Just as ChatGPT was trained on a wide range of internet text to produce a generalist chatbot that can respond to diverse text or spoken queries, medical LLMs are trained on clinical data contained in EHRs and other resources with the goal of producing 'foundation' models that a clinician can interrogate for medical insights. At Roche, Fabian Schmich is developing LLM-based digital twins that can predict a person's likely disease trajectory or response to treatment using deidentified EHRs

News in brief

# FDA go-ahead to test cellular rejuvenation therapy in humans

The first therapy directed at reversing diseases of aging by cellular rejuvenation is about to start clinical trials. Life Biosciences received clearance from the US Food and Drug Administration for a phase 1 trial to test their partial epigenetic reprogramming approach in people with eye disease.

Life Biosciences' ER-100 is a gene therapy that uses a modified adeno-associated virus vector to deliver genes encoding transcription factors OCT-4, SOX-2 and KLF-4 (three of the four so-called Yamanaka factors used to reprogram adult cells into stem cells). By applying controlled expression of the three genes, the company aims to restore methylation patterns and test whether partial cellular reprogramming can reverse aged or damaged cells to a younger state.

The upcoming trial will test ER-100 in patients with vision loss from either open-angle glaucoma or non-arteritic anterior ischemic optic neuropathy, two types of age-related vision loss caused by damaged ganglion cells. The trial's primary goal is to test ER-100's safety in 18 patients for up to five years; each patient will receive either a single high or low dose of the drug injected into one eye. ER-100 uses a doxycycline-inducible system, meaning that treatment can be halted if needed. Secondary endpoints will assess whether patients' vision improves.

Life Biosciences received the regulator's go-ahead for ER-100 after submitting an Investigational New Drug application containing preclinical and safety data obtained by co-founder David Sinclair of Harvard Medical School. Gene therapy *with Oct4*, *Sox2* and *Klf4* fully restored vision loss in injured mouse retinal ganglion cell models, as well as in old mice.

Life Biosciences is also investigating partial epigenetic reprogramming for liver disease with ER-300. The corresponding gene therapy with *Oct4*, *Sox2* and *Klf4* improved signs and symptoms of metabolic dysfunction-associated steatohepatitis in a mouse model.

including a vast cohort from Flatiron Health, a medical technology company acquired by Roche in 2018. "Since then, we have been exploring what can we do with this amazing data source of hundreds of thousands of patient records," says Schmich. In a publication from 2025, he and colleagues described a platform called DT-GPT and showed that it could outperform existing methods for forecasting changes in important physiological parameters in patients with lung cancer and Alzheimer's disease over weeks or even years.

Many of these first-generation medical forecasting models, including Delphi-2M, have focused on the structured components of EHRs, relying on well-defined test measurements and standardized definitions like the World Health Organization's International Classification of Diseases (ICD)-10 framework. But LLMs — and other transformer-based AI architectures more generally — also work well with messier text such as clinician's notes. "Over 80% of the information contained within healthcare exists as unstructured information," says Dobson.

The medical informatics community is also enthusiastic about AI tools that pair health records with other types of information. For example, a team led by Cory McLean at Google Research recently posted a preprint in which they complemented health records from the All of Us project with genomic data. This allowed them to introduce 'polygenic risk scores', which calculate the likelihood of developing particular medical conditions based on the combined effects of multiple disease-associated genetic variations. McLean notes that patients with more limited clinical data benefited the most from including genomic data, whereas in patients with extensive medical histories, it added less to the information clinicians already have. But for people with scantily documented medical backgrounds or those with a generally clean bill of health, McLean believes that genomic insights "could be particularly useful in early flagging for early or unexpected risk."

Meanwhile, the team behind Pheno.AI is taking human data collection to another level. As part of the Human Phenotype Project, Eran Segal's team at the Weizmann Institute has collected a staggering range of medical measurements from nearly 15,000 relatively healthy Israeli adults to build clinical trajectories and enable early signal detection. In addition to standard clinical tests, multi-omic analyses, body scans and microbiome samples, the cohort also integrates data on physical activity, diet, occupational questionnaires, medications and physical tests such as hand-grip strength.

Last summer, Segal and colleagues provided proof of concept for how these data could be used to build predictive foundation models for modeling metabolic and cardiovascular health and disease risk. "The main goal is to deeply phenotype a person, because we believe that's the only way that you can get into the insights that are hidden in the physiology," says Hagai Rossman, a former postdoc in Segal's lab who is now focusing on developing similar analysis tools for commercial deployment as chief scientist at Pheno.AI.

Some models, such as Delphi-2M, were designed to assess a person's risk of developing a wide range of various medical disorders well into the future. "We make an analogy with weather prediction, where we can forecast, and then basically we're going to give you probabilities of rain and sun and stuff like that," says Ewan Birney, executive director of the European Molecular Biology Laboratory and one of the lead authors on the Delphi-2M paper. Other efforts are focused on smaller, more modular AI models that simulate specific aspects of physiology in an effort to anticipate outcomes such as disease progression and drug response for that particular condition. For example, Schmich says that his team at Roche is working on pairing the healthcare data from Flatiron with additional data from cancer genomics company Foundation Medicine — another company acquired by Roche — to develop digital twins that could eventually help guide care for cancer patients.

The UK's Foresight-England project, developed by Dobson and colleague Chris Tomlinson, a clinician at University College London, is a national-scale effort to model short- and long-term health consequences of COVID-19. Although their project is formally focused on a single infectious disease, it will cover the full gamut of downstream outcomes that can result from this disease. "We're really asking sort of high-level questions about: given the patient's medical history, what will happen next?" says Tomlinson. "That gives us the ability to predict across a range of about 40,000 different events, from things like diagnoses, procedures, hospital admissions et cetera."

Like weather prediction, the confidence of the predictions generated by these models goes down as timelines extend further into the future. In his team's DT-GPT paper, Schmich found that predictions started to get noisy beyond a year, making it harder to derive confident interpretations. For Delphi-2M, Birney and colleagues devised a transformer-based

architecture that employs an attention mechanism that can capture timing-related patterns in medical histories. This enabled their model to identify sequential chains of medical events that are ultimately likely to result in a future diagnosis — for instance, marked changes in weight or blood glucose and insulin levels can anticipate onset of type 2 diabetes. Although their model could project decades into the future, long-range forecasts were inevitably more ambiguous and lower-confidence than those from just a few years out.

Many challenges await those keen to deploy these AI-based clinical predictors. Biomedical data are extremely personal, and AI model developers must take care to insulate this sensitive information. Shah and colleagues have grappled with this issue in developing Curiosity, a model trained on their Cosmos EHR dataset for the simulation of future patient health outcomes. As a solution, they only provide researchers with access to the model, which contains no identifiable personal information. "You can't take the data out, you can't take the model out," says Shah. Consent can also be a challenge. McLean says his team at Google has struggled to gain permission to access wearable-based data for model training, for example, and the Foresight-England initiative is on hold due to concerns raised by UK-based physicians' associations last spring about the reuse of COVID-19 patient data for AI training.

It is also unclear how big a dataset is needed to deliver useful, reliable forecasts. Birney estimates that data from at least half a million people would be needed to educate a model like Delphi-2M to make 'zero-shot' predictions — meaning the ability to deliver insights about medical conditions that it was not explicitly trained on. Training LLMs with bigger datasets yields better returns, but this is also expensive, and it can be hard to know where to draw the line. "They're compute-intensive, they're time-intensive, and we need to get them to a place where they're very accurate," says Vicki Seyfert-Margolis, founder and CEO of RespondHealth. "We've spent a lot of time on benchmarking to figure out what model we can use that's feasible in time and costs." Once that is achieved, says Seyfert-Margolis, the next step is to refine, check "and verify that it's accurate and correct."

The models also need to be trained on sufficiently diverse data to be broadly applicable across populations. Birney was pleased that Delphi-2M performed well at modeling disease risk in a cohort of Danish patients despite being trained on British data, but says it would be inappropriate to repurpose such a model for real-world applications without appropriate retraining. At Pheno.AI, Rossman says that his research team is focused on generating rich phenotype data from datasets in Japan and the Middle East to complement their existing Israeli dataset and produce forecasting tools that can better account for human genetic and geographic diversity.

When it comes to evaluating a future AI-generated disease forecast, there is also the fundamental issue of trust and transparency: how did the model arrive at its diagnoses? Clinicians draw on their extensive education and experience when interpreting a patient's test results, but many AI models have been described as 'black box' systems that simply spit out an answer without showing their work. "There's still a lot of nervousness around explainability," says Dobson. The LLM community in general has been working towards including so-called 'reasoning' capabilities that walk users through the steps leading to predictions and provide sources for relevant data. Schmich says his team is exploring such capabilities at Roche, but adds that "it's very hard to get ground-truth data for reasoning, unfortunately, without major efforts of clinical experts writing these down."

As a solution to this problem, RespondHealth has used a 'knowledge graph'-based approach for EHR analysis, in which formal medical principles and relationships between entities like genes, diseases and drugs are codified in a structured fashion on the basis of expert guidance. When they use LLMs and other transformer-based AI approaches to analyze patient EHRs and make diagnoses, this knowledge graph serves as a trustworthy reference and a 'guardrail' that ensures that the model's findings are consistent with medical reality. In parallel, the company also applies algorithmic methods that quantify the confidence of each prediction. "You don't have to necessarily trust the model," explains CSO Edward Kim. "We give you the receipt."

Personalized health prediction is probably the flashiest use-case — but also the farthest from reality at the moment, given the high stakes and regulatory uncertainty around clinical AI. However, there are other contexts where these tools could prove powerful. "We think that healthcare system modeling and prediction and resourcing is probably a nearer-term application," says Birney. "How many MRI machines do I need in the next two years in this location? How much service do I need to provision here?" Shah also sees this as an application for Epic's Curiosity model, allowing hospitals to maximize their

**News in brief**

# FDA relaxes oversight for wellness devices

The US Food and Drug Administration (FDA) announced on 6 January that it is loosening restrictions on wearables, including devices that measure blood glucose, blood pressure and other AI-enabled tools. The changes aim to spur innovation for digital wellness products and clarify previous guidance from the agency.

FDA commissioner Marty Makary announced the changes at the Consumer Electronics Show in Las Vegas. Previous guidance "didn't always make a lot of sense from the outside," he remarked in his speech. It had "unintended consequences" that "forced" developers to "build dumber software," he said.

The FDA's changes are described in two guidance documents released on the same day. The General Wellness: Policy for Low Risk Devices document greatly expands the scope of products that the FDA will consider 'wellness' devices that do not require regulatory oversight. The intention is to demarcate "clear lanes" for a consumer-grade versus a medical-grade device, Makary remarked at the conference.

The Clinical Decision Support Software document makes significant changes to how the FDA regulates tools that provide decision support for healthcare providers. Most notably, the changes say the regulator can, on a discretionary basis, exempt software that makes a sole recommendation (as opposed to giving providers a list of options), but it must meet certain criteria related to timing sensitivity. For example, software that predicts a patient's risk of future medical events could be exempt from FDA regulations, but if it predicts events that might occur soon — such as within 24 hours — it would be considered a medical device that triggers FDA oversight.

The FDA is also developing a "new regulatory framework for AI" that is "smarter and more forward thinking," Makary said in his remarks. At press time, that framework has not yet been announced.

News in brief

# VCs unite to spur Chinese academic innovation

A new venture capital (VC) company is raising its initial fund to spur a new generation of startups in China. Shanghai-based Apuri BioVenture believes that funding early-stage discoveries at universities is key to China's growing biotech sector. The intention is to move beyond China's early copycat approach to begin translating local innovation into startups. "China is entering a phase of the [scientific] discovery dividend," Apuri's founder Jun Bao told *BioCentury* in a recent interview.

Bao is a seasoned biotech executive who most recently was chief business officer of Biotheus, a bispecific antibody company bought by BioNTech in 2025. To form Apuri, he was joined by Frank Ye, a VC and former investment banker, and Jian Zhang, professor of stem cell biology at the Chinese Academy of Sciences and Fudan University, and CEO of two companies. Apuri's strategy is to source promising research from its network of scientists across China, looking for disruptive technology that has a plausible route to commercialization. Apuri builds relationships with academics early in their careers and remains engaged even when their research is not ready to spin out.

To boost a nascent company's seed-raising potential, Apuri has a hands-on approach that involves helping with financing plans, assembling a team, and creating intellectual property and business development strategies. In addition, Apuri prefers giving an academic researcher a CSO role, believing the demands of a CEO role to be too onerous when combined with a university position.

Apuri already has a portfolio of around ten emerging startups in its debut fund. The list includes a cancer diagnostics startup that uses a single blood draw to measure genomic histone modifications and infer originating tissue and stage. Others in Apuri's portfolio are in stealth mode, such as a regenerative medicine company developing embryoids from reprogrammed somatic cells and a Fudan University spinout that is targeting brown adipose tissue as a potential obesity treatment.

efficiency by forecasting patient stay durations, for example.

Modeling patient health trajectories and response to therapy could also be useful for posing 'what if' questions in the context of drug development and deployment. In clinical trials, for example, Schmich sees opportunities to use digital twins to identify inclusion or exclusion criteria, or to anticipate responses once a study is underway. "You could do these interim trial assessments, so while the trial is running, you are always one step ahead," he says. "Maybe you can forecast adverse events and manage them better for the patient's benefit." Seyfert-Margolis says that one of the use cases being explored at RespondHealth entails modeling the spectrum of patient response to various categories of drugs, and using those data to guide their application and identify underserved patient subgroups who aren't responding effectively. Since early 2025, the company has been collaborating with Microsoft to test how well their AI models can guide the use of the various blockbuster GLP-1 agonist drugs now on the market.

Even if it remains uncertain how these prototypes will ultimately evolve into clinical products, Birney is enthusiastic about exploring the opportunities that AI-powered medical modeling and forecasting might create in the future. "I'm getting the same feeling that I got 25 years ago with genomics," he says. "There's some very old problems where we've now got a completely new technique to tackle them."

**Michael Eisenstein**
Philadelphia, PA, USA