ADVANCING PRECISION HEALTH AND MEDICINE THROUGH
ARTIFICIAL INTELLIGENCE AND BIOMEDICAL DATA

STANFORD DEPARTMENT OF BIOMEDICAL DATA SCIENCE

ANNUAL REPORT 2025

Stanford MEDICINE | Department of Biomedical Data Science

CONTENTS

"As we enter a new era of biomedical discovery, the Department of Biomedical Data Science exemplifies how cross-disciplinary innovation can turn complex data into transformative insights. DBDS is redefining what's possible for science, medicine, and human health."

Lloyd B. Minor, MD
Carl and Elizabeth Naumann Dean of the Stanford School
Vice President for Medical Affairs Stanford University

Stanford's Department of Biomedical Data Science (DBDS) is contributing new approaches to fully harness the power of AI and data-driven discovery in academic medicine. While traditional academic structures often silo disciplines like statistics, biomedical informatics, and computer science, DBDS unifies them in one department. This structure sparks fundamentally new ways of thinking by seamlessly fostering deep collaboration — even within quantitative disciplines.

DBDS is grounded in two core values. First, we require biomedical analytical depth through "technical virtuosity." Second, we emphasize bilingual fluency — the ability to navigate both analytical rigor and biomedical domain expertise. Our graduate training program, supported by the National Institutes of Health (NIH) National Library of Medicine (NLM) for more than 40 years, instills methodological rigor and a focus on real-world biomedical impact. We train leaders who work at the intersection of data science and biomedicine, where we see data science as a melting pot of informatics, statistics, and computer science.

Now celebrating our 10-year anniversary as a department, we honor our deeper roots at Stanford. Four decades ago, early pioneers in the Stanford School of Medicine recognized the promise of clinical data long before electronic health records were standard — and, more than six decades ago, the need for rigorous and reproducible clinical trial research before biostatistics was a widely-recognized academic discipline. In the past decade, we've moved from single-modality approaches to integrating multimodal data across genomics, imaging, clinical records, laboratory tests, and patient outcomes.

DBDS is a department at the nexus of conceptual and impactful transformation.

As highlighted throughout this report, our faculty and students are developing transformative tools — including foundation models and generative AI — to advance our understanding of human health and improve care in unprecedented ways. Continued investment in this research, at this moment, has never been more important.

DBDS is a department at the nexus of conceptual and impactful transformation — at a significant time and place — embedded within the Stanford School of Medicine. We are a hub where clinical and basic science researchers come to pursue AI solutions to distinct questions and where data scientists come to engage with meaningful, real-world challenges.

Looking ahead to the next 10 years, we aim to continue to lead a technology-driven biomedical revolution by creating new ways to understand underlying human biology and transform clinical care in pursuit of a healthier future for all.

Sylvia K. Plevritis, PhD
William M. Hume Professor in the School of Medicine
Chair, Department of Biomedical Data Science

DBDS BY THE NUMBERS

50 FACULTY

54 PHD STUDENTS

90 MS STUDENTS

18 POSTDOCTORAL FELLOWS

40 RESEARCH SCIENTISTS

17 ADMINISTRATIVE STAFF

DEPARTMENT HIGHLIGHTS

Hosted the multi-sector AI Frontiers of Healthcare and Medicine Summit

Led a Stanford Medicine effort to hire AI faculty across disciplines and departments

Assistant Professor **Emily Alsentzer** was recognized by the Symposium on AI for Learning Health Systems for her contributions to this professional society

Professor **Trevor Hastie** won the 2025 C.R. and Bhargavi Rao Prize

Associate Professor **James Zou** won the 2025 International Society for Computational Biology Overton Prize

Student **Kristy Carpenter** won the NIH National Institute on Drug Abuse Drug Repurposing Challenge

Student **Oana Enache** was awarded the Stanford Interdisciplinary Graduate Fellowship

Student **Selina Pi** was a Lee B. Lusted Student Prize Competition Finalist

Student **Christine Yeh** won the Keystone Symposia Future of Science Award

GRADUATE STUDENTS

15 NIH NLM TRAINEES

8 NATIONAL SCIENCE FOUNDATION SCHOLARSHIPS

7 STANFORD GRADUATE FELLOWSHIPS

2 FULBRIGHT SCHOLARS

4 KNIGHT-HENNESSY SCHOLARS

2 STANFORD DATA SCIENCE SCHOLARS

4 WARREN ALPERT SCHOLARS

2 ARC INSTITUTE FELLOWS

2 RUTH L. KIRSCHSTEIN NATIONAL RESEARCH SERVICE AWARD INDIVIDUAL PREDOCTORAL F31 FELLOWSHIPS

1 SINGAPORE'S AGENCY FOR SCIENCE, TECHNOLOGY AND RESEARCH (A*STAR) SCHOLAR

1 STANFORD INTERDISCIPLINARY GRADUATE FELLOW

1 PAUL & DAISY SOROS FELLOW

LEADING THE BIOMEDICAL SCIENCE DATA ECOSYSTEM

We are at a pivotal moment in precision health and medicine. Through innovation, leadership, and partnership, DBDS is advancing a biomedical data ecosystem grounded in trust, transparency, and patient-centered outcomes. Join us as we shape this future.

- **Multimodal Foundation Models:** By integrating imaging, genomics, electronic health records, and more, DBDS is laying the groundwork for models that reflect the full complexity of human health.

- **AI Agents for Healthcare:** DBDS is driving the development of agentic AI systems — tools that support clinical decisions with autonomy and explainability.

- **Synthetic Data Innovation:** DBDS is reimagining clinical trials using real world-informed synthetic data to create more inclusive, efficient, and ethical designs, especially when real control arms are not practical or feasible.

- **Warren Alpert Foundation Grant for AI and Computational Biology:** This generous support is shaping the future by enhancing training and retention of scholars in computational biology and AI.

- **Cross-Sector Leadership:** DBDS convenes experts across academia, biopharma, healthcare, and technology to tackle AI's most pressing challenges in health. Strategic relationships with Accenture, Microsoft, Google, and others deepen our reach and impact.

- **Collaboration & Careers:** This dynamic annual platform fosters collaboration and career opportunities by uniting top researchers, students, and industry leaders.

MORE THAN 40 YEARS OF

"AI has the potential to revolutionize medicine, tailoring treatments based on genetic, environmental, and lifestyle factors, much as algorithmic trading transforms large datasets into investment strategies."

MARTY CHAVEZ ('90)
American financier and entrepreneur, and former Chief Financial Officer and Chief Information Officer of Goldman Sachs.

"DBDS graduates are advancing the field of precision oncology. By applying AI/ML technologies, we're accelerating drug discovery and optimizing clinical development to bring effective cancer treatments to patients more efficiently."

SHIRLEY LIU ('02)
Founder and Chief Executive Officer of GV20 Therapeutics

"The depth of knowledge and breadth of research across DBDS faculty puts trainees at the forefront of harnessing AI to improve clinical decision support safely and reliably."

JESSIE TENENBAUM ('07)
Associate Professor of Biostatistics and Bioinformatics, Duke University

WORLDCLASS TRAINING

"The strength of DBDS is bringing together AI experts, biologists, and clinicians to fully realize AI's potential in medicine. Applying AI to complex datasets will help decode biology's complexity and identify patterns impossible for humans to detect."

SERGE SAXONOV ('06)
Co-founder of 10x Genomics that pioneered single-cell and spatial genomics

"DBDS provided me with the technical expertise to develop comprehensive multimodal AI models, while keeping me firmly grounded in what truly matters. My life's purpose is creating technology that makes a meaningful difference in people's health, complementing theoretical advances with real-world impact."

JUAN MANUEL ZAMBRANO CHAVES (MS '20, PHD '24)
Senior Researcher, Microsoft Research Health Futures

"My DBDS training immersed me in a highly interdisciplinary environment that prepared me exceptionally well for my current work. I use data-driven approaches across genomics, epidemiology, and clinical research to develop more personalized, long-term care strategies to improve health outcomes for survivors of childhood cancer."
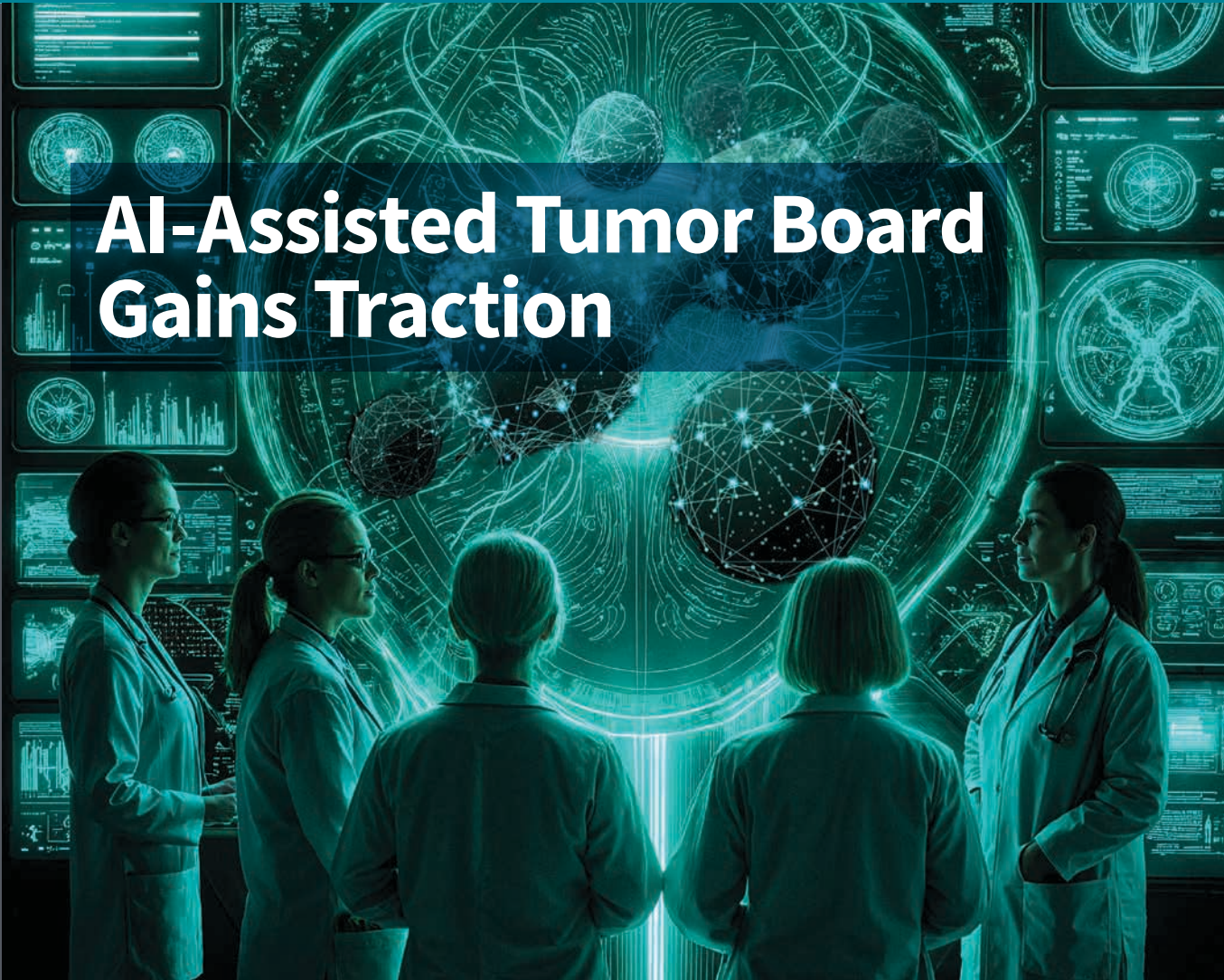
TIFFANY EULALIO ('24)
Postdoctoral researcher, St. Jude Children's Research Hospital

DBDS is pioneering AI-assisted tools, like tumor boards, as well as informing the use of large language models (LLMs) in health systems and patient care.

By integrating multimodal data, addressing LLM limitations, and ensuring real-world evaluation, DBDS faculty are advancing safe, effective AI that supports — not replaces — clinical judgment in complex medical decision making.

AI-Assisted Tumor Board Gains Traction

A Stanford Medicine team has made significant progress over the past year developing an AI-assisted multidisciplinary tumor board. It combines different types of cancer data such as images, clinical notes, laboratory values, and genomic tests, and then quickly extracts patterns or trends to guide precision cancer care.

Cancer care involves complex decisions for patients, families, and medical professionals. To help work through decisions and devise treatment plans, tumor boards bring together specialists including oncologists, surgeons, and other doctors; nurse case managers; and administrators. Despite their importance, however, tumor boards face challenges. They rely on complex data, fragmented information across health systems, and limited time for in-depth discussion about individual cases.

Although use of AI in clinical decision-making is still not reliable enough for routine clinical use, several AI features may make it a helpful assistant to human-run tumor boards. With support from the Advanced Research Projects Agency for Health (ARPA-H) and the leading global professional services company Accenture, a Stanford Medicine team has made significant progress over the past year developing an AI-assisted multidisciplinary tumor board. It combines different types of cancer data such as images, clinical notes,

electronic health record foundation models that integrate clinical text and structured data, using data-centric methods to guide how the models are trained.

Assistant Professor Akshay Chaudhari is developing a system that learns from radiology images without needing labels (called self-supervised learning), while Assistant Professor Roxana Daneshjou is creating an AI-assisted data-labeling tool. To house all the data for analysis by Stanford scientists, Stanford Medicine's Chief Technology Officer Todd Ferris is building

Foundation models are fast and versatile AI tools that can be adapted to do many different tumor-board related tasks such as summarizing text, analyzing medical data, and even answering questions in real time.

laboratory values, and genomic tests, and then quickly extracts patterns or trends to guide precision cancer care.

Foundation models are fast and versatile AI tools that can be adapted to do many different tumor board-related tasks such as summarizing text, analyzing medical data, and even answering questions in real time. Because foundation models do not need to be newly designed for every clinical task, they can be reused for a range of customized applications.

Building an AI-assisted tumor board is a large and complex undertaking, involving many different types of expertise to transform patient data into customized knowledge for use by a human-powered tumor board.

A multidisciplinary team at Stanford, led by Professor Sylvia Plevritis and with funding from ARPA-H's Biomedical Fabric Toolbox, is tackling these challenges. Their efforts are focused on developing "patients like me" tools to search and analyze large medical databases of clinical records and images to find patients with similar clinical features, treatment histories, genetic profiles, and outcomes. These tools can guide precise treatment decisions, better match patients to clinical trials, and improve insights about how a patient's disease may progress.

Professor Nigam Shah and Research Engineer Jason Fries are developing

a secure, cloud-based system that stores de-identified cancer data linked to patients.

Professor Rob Tibshirani's task is to figure out how to use AI to enable all the datasets to talk to each other in the same language. His team has developed a "cooperative learning" method that automatically brokers relationships between different data types and learns when they agree on predictions about patient outcomes.
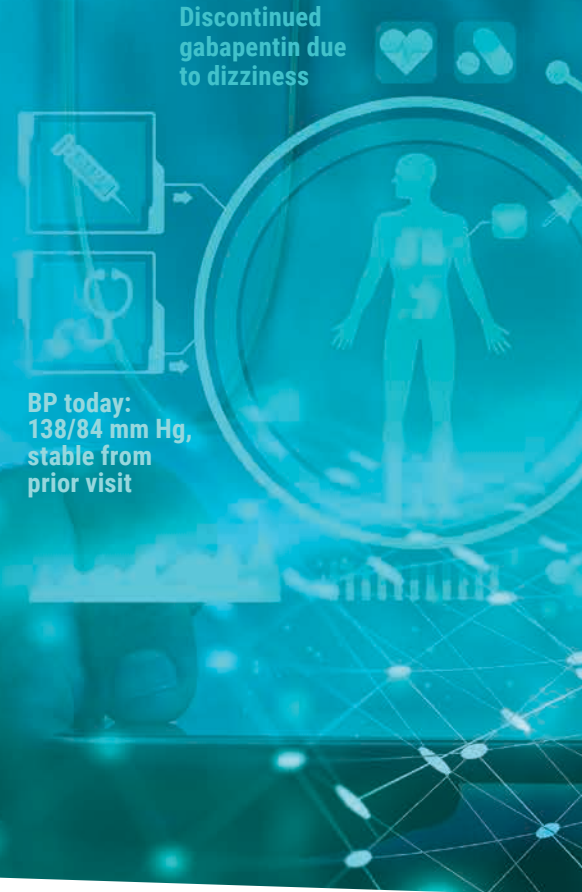
What does all this look like? Assistant Professor Manuel Rivas is making a data dashboard that summarizes all of a patient's medical data into a customizable, user-friendly display of laboratory values, medical history, prognosis, and more clinical information that can help decision making about treatment.

A key part of this exciting and ambitious project will be determining how well an AI-assisted tumor board works. Executive Director Karen Matthys and Professor Emeritus Daniel Rubin are learning more through a landscape analysis of process, pain points, unmet needs, and opportunities for improvement.

These insights will be critical to guide development, testing, and ultimately, implementation of precision oncology resources like AI-assisted tumor boards. Their success could dramatically improve care for patients and their medical teams aiming to provide the best possible personalized care.

Discontinued gabapentin due to dizziness

BP today: 138/84 mm Hg, stable from prior visit

LARGE LANGUAGE MODELS HAVE ARRIVED:
Is Medicine Ready?

More research is needed to continue to develop and implement powerful AI tools accurately, safely, and without bias.

There is an ever-increasing presence of AI in patient care. Are large language models (LLMs) up to the task? Yes and no. LLMs are one type of AI that focuses on human language, drawing from text sources across the internet to answer questions, summarize documents, and write stories after being given a specific task, or prompt. LLMs can operate through "chatbots" like GPT-4, Gemini, and Claude, learning from patterns. They are very good at analyzing and interpreting massive amounts of written material — from multiple sources ... in seconds.

But LLMs only know what they have been told, which is the data used to educate, or train them. They are not human and can be out-of-date depending on the last thing they were taught. LLMs can also generate inaccurate information, or "hallucinate" — without knowing they

are wrong. These shortcomings can pose a problem for patient care.

New work from Assistant Professor Roxana Daneshjou shows that LLMs don't perform well on tasks such as nuanced back-and-forth conversations that occur between doctors and patients. She developed a framework[1] that uses simulated doctor-patient conversations between AI "agents" to test LLM performance on clinical reasoning, history-taking, and diagnostic accuracy. The findings reveal key limitations, and she offers recommendations to improve the diagnostic performance and safety of off-the shelf LLMs that are not trained on actual patient data.

Writing recently[2] in the *New England Journal of Medicine — Artificial Intelligence*, Daneshjou and Assistant Professor Emily Alsentzer argue that medical licensing exams commonly used to evaluate LLMs give a limited and

MRI revealed
chronic micro-
vascular ischemic
changes

50%

Decreased
FEV1 noted on
spirometry

75%

25%

ER+/PR+/HER2-

incomplete view of how useful the models perform in real-world care. Because such exams were never designed to reflect the complexity of real-world clinical reasoning or care delivery, Alsentzer and Daneshjou call for more meaningful evaluations of LLMs grounded in real-world clinical workflows.

Alsentzer also warns that LLMs trained mainly by text from the internet can present serious challenges when applied to clinical tasks. In a recent study,[3] her team tested the chatbot GPT-4 in four medical scenarios: writing clinical education materials, helping doctors with diagnoses, making care recommendations, and evaluating patient symptoms. The results were concerning: GPT-4 often reproduced or amplified common stereotypes. For example, GPT-4 was less likely

to suggest advanced imaging scans for Black patients, even when their symptoms were the same as individuals who were not Black.

However, there are some bright lights. For example, new research from Assistant Professor Akshay Chaudhari shows that LLMs are useful for document-related health-care tasks like summarizing radiology scan findings, describing a patient visit, or determining problem lists from clinical notes. He and his team put to the test 8 LLMs across 12 medical specialties: How well could each complete these tasks compared to a real doctor? In most cases, the LLMs spit out summaries that were as good as or better than[4] those written by medical experts — potentially easing time-consuming administrative tasks.

"Fast thinking" AI tools are also getting better at knowing how to determine a diagnosis and decide what to do next, a skill known as clinical reasoning. Assistant Professor Jonathan Chen, Faculty Director for Medical Education in AI, conducted one of the first rigorous clinical trials to evaluate LLMs with real doctors making real diagnoses.[5] He and his team were very surprised that GPT-4 alone itself outperformed doctors (with or without AI help).

Clearly, more research is needed to continue to develop and implement powerful AI tools accurately, safely, and without bias. What's more, while AI has enormous potential to revolutionize medicine, it can't do that without hands-on support from humans, including training models with real-world clinical data.

Keeping a Pulse on LLMs in Health Care

Real-world usage of LLMs in healthcare systems may be outpacing their readiness for prime time, says Professor Nigam Shah, who is also Chief Data Scientist for Stanford Health Care. Toward verifying the accuracy of these models, Shah and his team are using several large, real-patient datasets (including Stanford's EHRSHOT,[6] INSPECT,[7] and MedAlign,[8] and many others) to develop a standardized way to test and compare the ability of six LLMs to perform real-world clinical tasks.

The new LLM-grading framework for healthcare builds on the Holistic Evaluation of Language Models (HELM)[9] developed by Stanford's Center for Research on Foundation Models,[10] and is being shared widely for further testing. HELM tests

Shah and his team are using several large, real-patient datasets to develop a standardized way to test and compare the ability of six LLMs to perform real-world clinical tasks.

LLMs various ways, including for accuracy, fairness, bias, robustness, and deployment. In a similar fashion, the newly developed MedHELM framework[11] grades LLMs on their ability to complete 121 clinical tasks in the environment in which they are being used. That may be in small rural hospitals, community clinics, or in much larger health systems. Evaluated tasks are grouped into five main categories: clinical decision support, clinical note-making,

patient communication/education, medical research support, and administration/workflow.

How did the LLMs do? Results varied significantly by model and healthcare system characteristics. Large models excelled at complex reasoning, such as medical calculations and detecting bias. Medium models did better on simpler tasks, such as predicting readmission risk. Small models handled basic tasks such as summarizing answers to consumer health questions but struggled to complete tasks requiring specialized medical or mental health knowledge. This research is an important contributor to global, responsible use of AI in healthcare — offering ongoing pulse checks on LLM performance in a range of different clinical settings.

Reading Reddit

Reddit is a popular online platform for advice, news, and support. Professor Tina Hernandez-Boussard sees Reddit as something else — a rich source of health data.
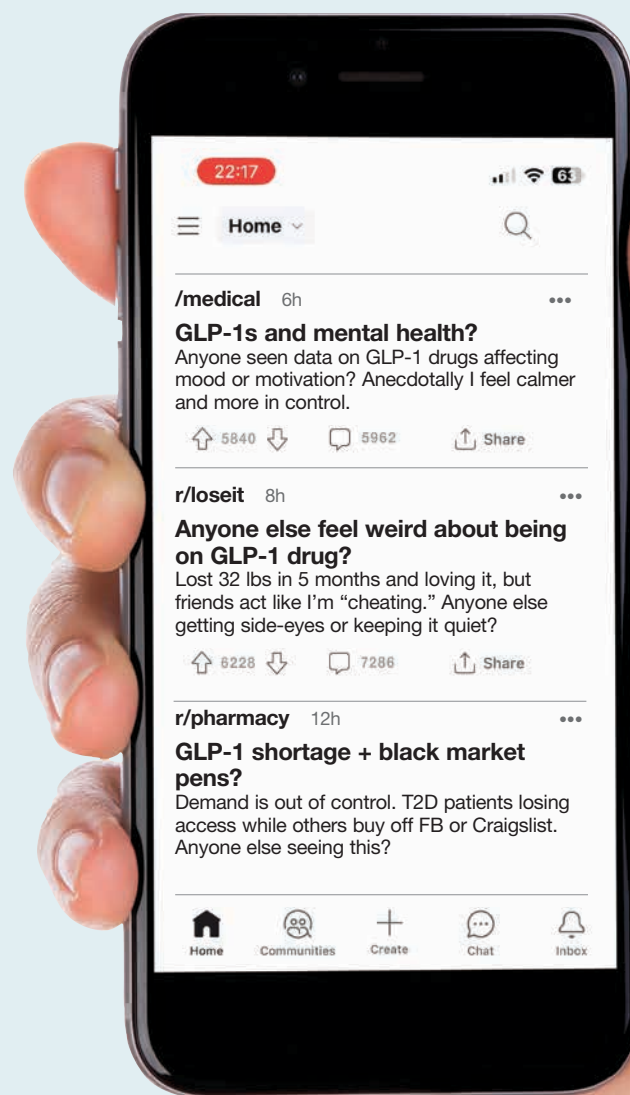
Typical sources of health data are electronic health records that document a patient's medical history, diagnoses, medications, test results, and treatment plans. But they do not describe personal views and concerns people have about their health. Some people, for example, do not trust doctors or their healthcare system and don't speak up during health visits about important issues that affect them.

Working with Stanford cardiologist Fatima Rodriguez, Hernandez-Boussard turned to Reddit feeds, which are public, to look for patient-centric, unfiltered information about GLP-1 medications for diabetes and obesity. GLP-1 medications tweak hormone signaling in the body that helps people feel full longer and eat less — and have transformed treatments of these chronic conditions.

She and her team used AI to sift through more than 390,000 unique Reddit posts[12] and automatically detect common topics like weight loss, side effects, or medication access. This approach allowed them to identify and group key themes in public discussions about the medications. The stepwise approach involved "cleaning up" the text, identifying topics, grouping them into themes, and summarizing results over and over.

What did they learn? In general, people were neutral to positive about GLP-1 medications — but access and affordability were key concerns. Hernandez-Boussard is repeating the Reddit approach to study various other health-related topics, such as well-being and stress experienced by caregivers for older people by using different sources of text from online discussion platforms.

Reddit posts in image (right) are fictitious.

DBDS scientists are unearthing biological insights, using creative ways to analyze and integrate data.

From tracking how cells age to reshaping genomic medicine, precision oncology, and cancer screening, DBDS faculty convert messy, multimodal data into actionable knowledge that will improve diagnostics, personalize treatment, and accelerate progress toward improving human health.

How Cells Grow Up in Health and Disease

As building blocks of the human body, cells are used in numerous ways to maintain health. Red blood cells carry oxygen, neurons send electrical signals, and immune cells fight infection. Cells change as they age — maturing to develop new abilities that help organs do their complicated work. New research illuminates this process toward both understanding health and fighting disease.

Associate Professor Aaron Newman developed CytoTRACE 2,[13] a new method that accurately measures how easily cells mature into 150 different human or mouse cell types. Unlike typical "black box" AI approaches, Newman used an interpretable deep learning model that not only identifies patterns from large amounts of data, but also explains why the model makes decisions, thus generating new knowledge.

The brute force effort of feeding detailed molecular data into the model took years but should be useful for various cell-based therapies related to cancer or organ repair and transplantation.

Focusing on the brain, Associate Professor James Zhou's laboratory and their collaborators created a detailed molecular map of how mouse brain cells change with age. He analyzed more than four million cells across different ages, after exercise, or after applying a molecular technique that reprograms cells to a younger state. Applying AI to create spatial aging clocks,[14] his team learned that some cells that are rare in the brain, such as immune (T) cells, exert a pro-aging influence on neighboring brain cells. This information may help explain how inflammation damage contributes to Alzheimer's disease and other aging-related brain disorders.

THE NEW GENOMIC MEDICINE

Imagine trying to create an enormous, searchable library from an unorganized collection of handwritten books. Each one has typos and notes written in the margins: all in different languages and styles. Some pages are missing, and others are repeated. Making this library useful requires translation, proofreading, and cross-referencing everything.

In a similar sense, a lot of careful, and often tedious, work is needed to organize, or curate, genomic data to make it readable, reliable, and searchable. That's because technologies to read genomes are not fool-proof, creating mistakes. It's also not immediately clear to scientists which of the many genetic changes across people are meaningful (as in contributing to disease) or merely an

Professor Klein is now putting data science to work, integrating pharmacogenomics, genomics, and disease knowledge.

alternate spelling that may have no tangible effect.

For years, the NIH-funded Clinical Genome Resource,[15] or ClinGen, has helped doctors and scientists understand which genes and genetic changes are linked to disease, using clear and reliable evidence. ClinGen works through a crowdsourced model of experts and working groups that focus on specific disease areas.

Professor Teri Klein, a long-standing expert in pharmacogenomics (how genes affect drug response), led efforts to add pharmacogenomic data into ClinGen, enriching its value for understanding both disease diagnosis and disease treatment. As part of the ClinGen Rheumatologic Autoimmune Disease Clinical Domain Working Group, Professor Klein is now putting data science to work,[16] integrating pharmacogenomics, genomics, and disease knowledge. The goal is to advance understanding of hard-to-treat autoimmune diseases such as lupus, rheumatoid arthritis, and gout.

A key deliverable from this work will be a clear system for organizing and reviewing genomic data, using standard methods to decide which genetic changes are important for autoimmune diseases.

Shaping Up Cell Therapy for Cancer

About 10 years ago, engineered immune therapy was approved by the U.S. Food and Drug Administration (FDA) to treat certain cancers of the blood, like leukemia and lymphoma. Chimeric antigen receptor T cell, or CAR-T, therapies have revolutionized cancer care for these conditions, personalizing treatment to an individual's specific cancer. In CAR-T therapy, doctors change a patient's immune (T) cells to better fight cancer. That includes helping T cells find tumors, as well as using gene-editing techniques to make the cells stronger and help them multiply faster. In many cases, CAR-T therapy means a single treatment instead of multiple infusions — potentially doubling or tripling the time individuals remain cancer-free.

But there are hurdles, including side effects. Personalized therapy is also expensive, since every individual's drug is made from scratch. A main current challenge is that CAR-T therapy does not work well for so-called solid tumors that reside in organs. That is because these tumors embed themselves within a biological fortress of many cell types and other materials that feed the tumor, supply it with oxygen, and protect it from immune attack. For now, CAR-T therapy is a last-option therapy for solid tumors.

New research[17] from Professor Barbara Engelhardt combined off-the-shelf, inexpensive video live-cell imaging with a new AI-powered method that tracks, learns, and measures modified T cell behavior in real time. Engelhardt also developed a tool to detect patterns of T cell properties and actions. For example, size — T cells tend to get bigger (more activated) after confronting a cancer cell. The tool also measures how well T cells recruit more of their own to help fight cancer, as well as how good each cell is at stopping cancer from growing. All of this information, Engelhardt predicts, could be put to use to design precision CAR-T therapies for hard-to-treat solid tumors.

A Better Way to Screen for
PROSTATE CANCER

No one wants to suspect they have cancer when they don't. But that's what happens to about 75% of men who have a prostate biopsy due to high blood levels of prostate-specific antigen, or PSA. The opposite also happens. PSA-guided prostate cancer screening can yield false negatives: missing those men who actually do have cancer.

Getting it right matters. Prostate cancer is still the most common cancer among men in the United States. But many PSA-detected hidden cancers are slow-growing and may not require immediate treatment. Due to these limitations, whether or not to use PSA levels as a screening method depends on a man's individual risk — such as being from a particular ethnic background or having a family history of prostate cancer.

New research from Professor John Witte and his team shows that some men have high levels of PSA unrelated to cancer risk: due instead to their genetic ancestry. To improve screening for the most aggressive types of prostate cancer, they developed a method to adjust PSA readings to account for a man's genetically based PSA levels.

The scientists conducted a large study[18] involving nearly 400,000 men from varying ancestries who did not have prostate cancer. The search turned up 450 genetic changes linked to PSA. Applying a statistical method to weed out those PSA changes unrelated to prostate cancer, Witte and his team came up with an accurate way to adjust PSA levels to reflect true cancer risk. This type of adjustment also better predicted risk of aggressive tumors that are best treated when found very early.

To improve screening for the most aggressive types of prostate cancer, they developed a method to adjust PSA readings to account for a man's genetically based PSA levels.
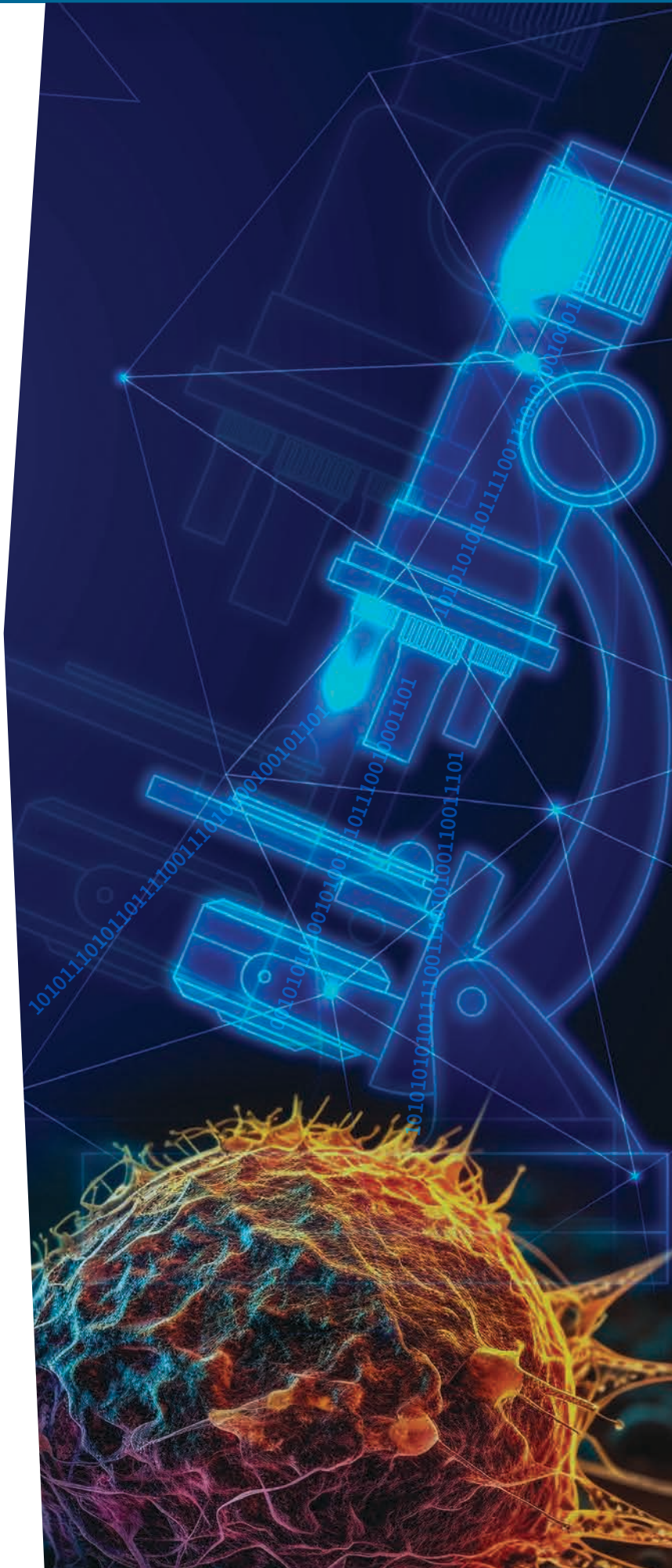
MODERN
MICROSCOPE

Pathology is one of the oldest branches of medicine. For many years, pathologists have studied disease by examining organs and tissues, using a microscope to carefully review samples on thin glass slides.

Today, pathology slides remain critical to understanding and treating disease, but new tools are available to analyze them. Many academic medical centers and independent laboratories in the United States have adopted digital pathology, in which glass slides are converted into digital images (whole-slide images, or WSIs).

These changes have ushered in a new world of computational pathology made possible through AI innovations. While human review is subjective and limited, AI-based methods can "see the unseen" in WSIs from tumor samples. New tools can analyze every cell on a slide, cancerous or not, to learn more. That means being able to predict gene activity linked to important cancer processes by using tissue images alone.

Associate Professor Olivier Gevaert has embraced this new world, recently unveiling a new AI-based linear transformer model[19] that predicts cancer gene activity from WSIs. Transformer AI models break information into small parts and learn how those parts relate to each other, similarly to how people follow a conversation. These models learn very rapidly from huge datasets, detecting patterns and connections that pathologists might overlook (or could never actually see in the first place).

Gevaert and his team trained their model with data from 7,500 tumor samples (from 16 cancer types) and proved that it also worked with separate datasets. The tool predicted which breast cancer patients might benefit from more aggressive treatment without the need for expensive genomic testing — a step on the path to precision oncology.

Who's Who in Cancer

The tumor microenvironment (TME) is a complex and dynamic mix of cells, molecules, and blood vessels surrounding and supporting a tumor. Over the years, research on the TME has helped advance cancer immunotherapy in which immune cells can be summoned to attack a tumor.

New research[20] from Professor Sylvia Plevritis details a powerful new way to track the ever-changing interactions and relationships between cell neighbors in the TME. Instead of just looking at what is inside a cell, Plevritis' spatial biology approach looks at how and where cells are arranged and work together. It can also look at TME changes in different scenarios, such as after treatment.

The new work debuts a method that displays how different cells are arranged in 3D laboratory-grown tissue models (assembloids) consisting of lung cancer cells and support cells from real human tumors. Assembloids mimic actual TME structure and activities.

Plevritis' method uses math and statistics to measure how often different types of cells in the TME appear near each other in intact tissue samples, adjusting results to enable fair comparisons across different samples. This method allows identification of biologically meaningful patterns of thousands of cell-cell arrangements across different types of TME data.

A key finding was a sinister role for scaffolding cells called fibroblasts that live in different neighborhoods of the TME. Plevritis and her team were also surprised to learn that some cancer treatments actually rearranged cell positions (although not cell types or amounts) to look like the original untreated tumor — suggesting a "spatial signature" for drug resistance and a target for novel therapies.

Innovative methods advance biomedical research in innumerable ways.

DBDS faculty are advancing statistical tools, AI-enhanced diagnostics, and patient-centered clinical trials. This research involves rigorous, transparent approaches to ensure that biomedical data reflect real-world health impacts — bringing clarity, accuracy, and humanity to data-driven science.
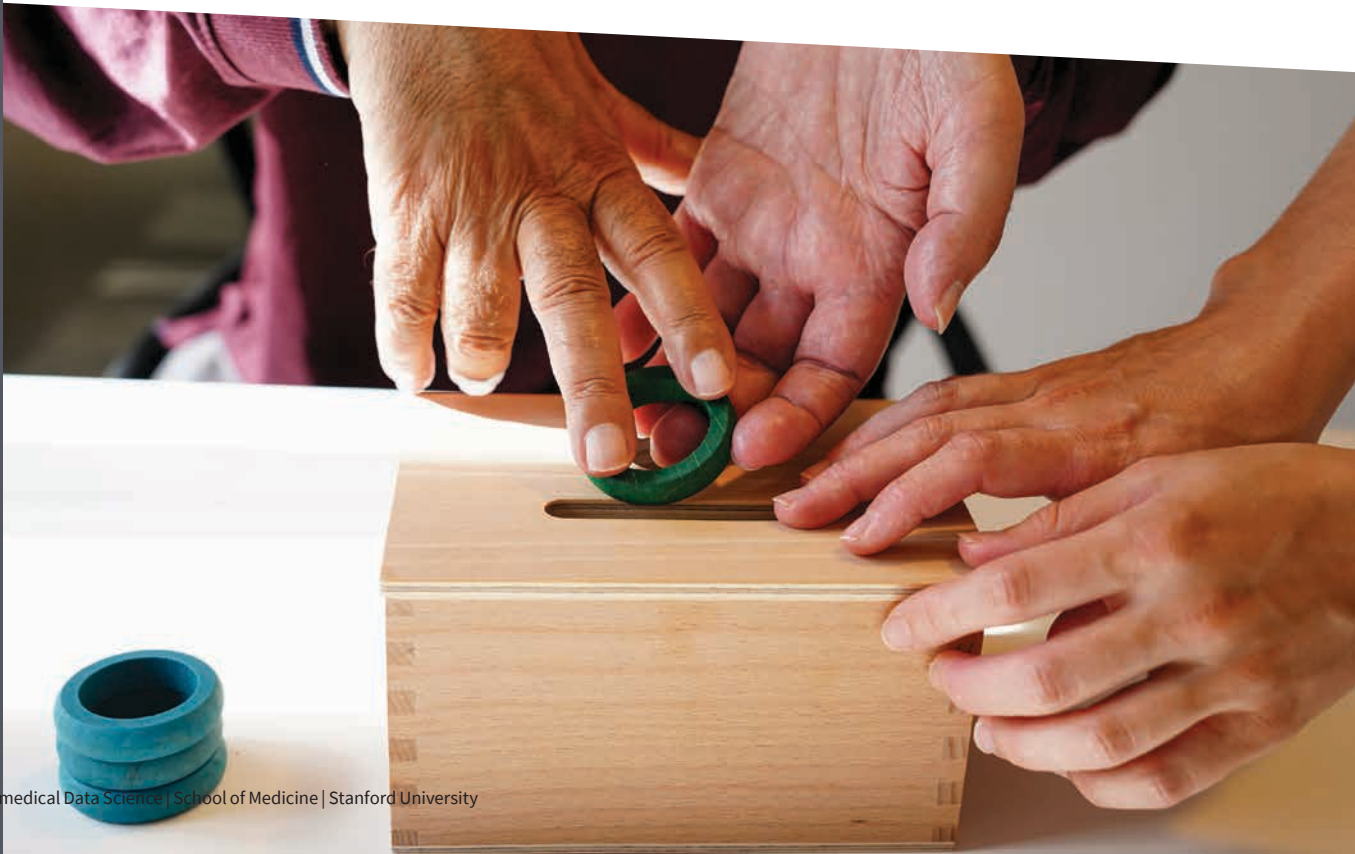
Putting Patients First

Clinical trials test whether medical treatments work, but other questions are also important. For example, which aspects of a treatment are most important to people living every day with a particular condition? Patient preference matters because treatments are most effective when they match what people want and need.

Professor Ying Lu develops clinical trial designs that consider what matters to patients. His new work[21] focuses on amyotrophic lateral sclerosis (ALS). Lu started with data from a year-long online survey of people with ALS in which individuals rated how well they could complete certain daily tasks (such as walking, speaking, or eating) and then ranked which tasks mattered most to them.

To tease apart the impact of these preferences on patient outcomes, Lu reviewed national health records to discover how long ALS patients lived. He analyzed survey responses in which patients ranked which functions mattered most to them and then used statistical comparisons to look for any effect on their lifespan. The results revealed that patient-reported priorities better predicted survival than traditionally used clinical measures.

Lu is now leading a Department of Defense-funded project to find better ways to measure what outcomes are most important to ALS patients. The project is conducting an international online survey and follow-up study to test new digital questionnaires that capture patient priorities. Overall, Lu aims to make clinical trials more efficient by incorporating what matters to patients into how clinical trials are designed at the very beginning.

LLM-Lasso: A Better Cancer Classifier

Lymphoma is a cancer of the immune system in which white blood cells called lymphocytes grow rapidly and form tumors. There are dozens of different lymphomas; using traditional pathology methods, it can be hard to tell them apart.

A new era of pathology is dawning, thanks to comprehensive and efficient new AI tools.

Professor Rob Tibshirani teamed up with Stanford computer scientists and oncologists to modernize a popular tool in statistics and machine learning he introduced 30 years ago, known as Lasso (short for Least Absolute Shrinkage and Sele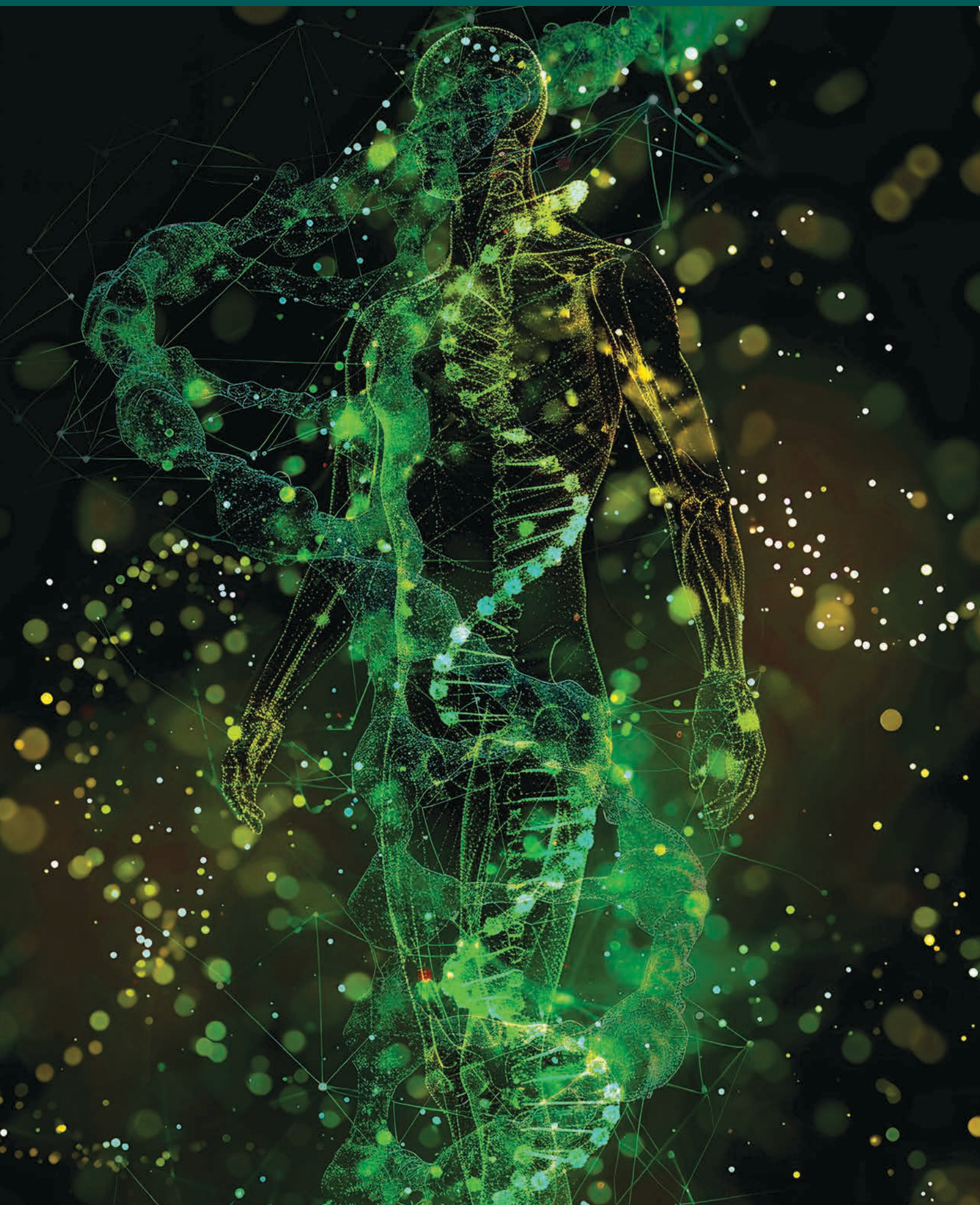ction Operator). They have created a new version called LLM-Lasso,[22] combining the original method with a large language model (LLM) to add AI-powered insights.

The scientists identified several key genes (AICDA, BCL2, and BCL6) that could help distinguish two similar types of lymphoma.

The team fed LLM-Lasso information from more than 100 lymphoma samples, including a list of about 1,500 specific genes and how active 500 of those genes were in each sample. They also provided a detailed description of the medical problem they wanted the model to help solve.

The LLM-Lasso model next incorporated lymphoma-related knowledge pulled from trusted scientific sources using a technique called retrieval-augmented generation, or RAG. Tibshirani's team helped guide the model by labeling which sources to trust and which to ignore, improving how the model selects the most important information based on both raw data and expert knowledge.

As a result, the scientists identified several key genes (AICDA, BCL2, and BCL6) that could help distinguish two similar types of lymphoma — follicular lymphoma and diffuse large B-cell lymphoma — potentially improving diagnosis and opening the door for new discovery.

SIMPLIFYING SCIENCE

Data is the heart of biomedical science, and modern methods are generating a lot of it. While only a few years ago large datasets were a hindrance, today's AI methods devour them to find new insights.

One of the most common types of health-related research that generates large, complicated datasets is an observational study, in which scientists watch what happens in the real world without interfering. Observational studies collect lots of data about people's health, behavior, or environment to look for patterns or connections.
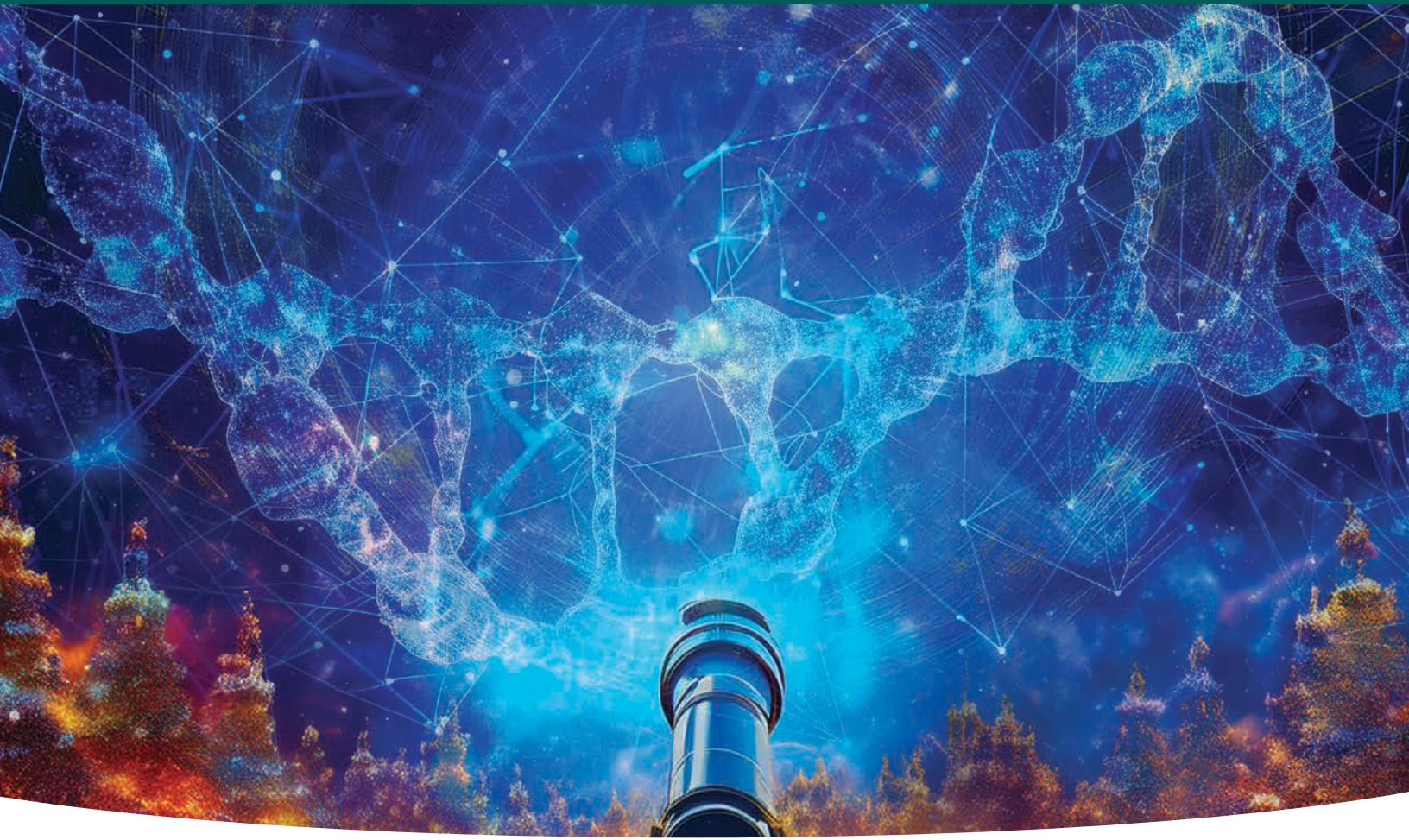
It's not hard to imagine how some of these values could interact to affect experimental conclusions; in research, it's a concept known as "covariates." Take skin cancer, for example — although genetic traits affect risk, so do many other factors like sun exposure and skin tone. Teasing apart covariates, and their respective or combined influences, is important for determining cause-and-effect relationships that guide prevention and treatment.

New research[23] from Professor Wing Wong provides a better way to study cause-and-effect relationships using complicated real-world data from observational studies. Wong developed a new method that combines deep learning (a type of machine learning) that excels at finding complex (nonlinear) patterns with a new modeling approach to better estimate cause-and-effect relationships. Wong and his team used the method to track how covariates in observational studies relate to each other in ways not easy to see. That can improve understanding of the impact of specific treatments on health outcomes, which may be useful for making healthcare decisions.

Separately, Professor Chiara Sabatti's recent work[24] might also simplify matters for scientists who conduct observational studies, in particular one type known as a genome-wide association study (GWAS). In these studies, researchers look at all the genes in many people's DNA to find out if certain genetic differences are more common in people with a specific trait or disease. Like other observational studies, GWAS can show associations but can't prove cause-and-effect relationships.

Sabatti aimed to separate true causes from those just linked by coincidence, at the highest possible resolution. She and her team developed a new statistical tool to weed through groups of DNA variants in very large datasets to find which ones are uniquely related to disease risk. One advantage of this technique is that it lets researchers test hypotheses carefully without relying on unrealistic assumptions (such as assuming variables are not related). Another plus is that the method works like a "wrapper" around machine learning methods — helping to explain the model's results but making sure discoveries are scientifically valid, not just patterns the model happened to find.

Computational Bio-Telescope
LOOKS BACK IN TIME

It's often said that understanding the future means studying the past, and that can be true in science and medicine. Professor Julia Palacios and her team have developed a "computational bio-telescope" that reconstructs how populations change over time,[25] analyzing DNA from ancient samples (in this case, from bison bones). Because these samples are often collected at different times and thus combine to make large, messy datasets, clever methods are needed to decode their history efficiently.

Building on a common statistical approach for tracing changes in population size, Palacios adapted a model called the "Tajima coalescent" that doesn't track every individual DNA sequence but instead groups them based on when they were collected. Her team could thus deduce historical population trends from genetic data more easily and efficiently.

Palacios' new approach combines advanced statistical methods with powerful computational tools, offering a better way to study evolution over time. Although this particular example focused on ancient bison DNA, the research strategy has broad potential applications. Those include tracking how viruses evolve, guiding vaccine development, and improving genomic research for personalized medicine.

A Fairer Way to Judge New Medications

Developing new medications is a complex process. Once a potential drug is found to be safe in humans, large clinical trials determine whether it achieves its intended effects — such as shrinking tumors, relieving pain, or many other health outcomes.

Another key measure is duration of response (DoR): how long a treatment benefit lasts after a patient first responds. DoR is especially important when overall survival or cure is difficult to achieve.

Imagine a new cancer medication with an average DoR of 20 months: meaning that, on average, patients receive nearly two years of

Lu Tian developed a more informative method for measuring DoR, provides a more accurate and honest picture of how a new treatment works across all patients.

treatment benefit. Even among those individuals, the duration varies — some benefit for more than a year, and others relapse within weeks. As well, the averaged DoR does not take into account patients who don't respond at all, meaning it's not completely accurate. For example, a drug that benefits very few patients for a long time might appear more effective than one helping many patients for a shorter period, even if

the latter offers greater overall value.

To address this problem, Professor Lu Tian developed a more informative method[26] for measuring DoR. Instead of not including patients who didn't respond at all in the DoR calculation, his method counts them as having a zero-response time and estimates how long each response lasted, even when some data is missing or incomplete. This strategy provides a more accurate and honest picture of how a new treatment works across all patients (not just those who respond), offering a clearer understanding of a new medication's real-world impact.

PRIMARY FACULTY | DEPARTMENT OF BIOMEDICAL DATA SCIENCE

SYLVIA K. PLEVRITIS, PhD
Professor & Chair

EMILY ALSENTZER, PhD
Assistant Professor

ROXANA DANESHJOU, MD, PhD
Assistant Professor

BARBARA ENGELHARDT, PhD
Professor

TREVOR HASTIE, PhD
Professor

ALEX IOANNIDIS, PhD, MPhil
Assistant Professor

IAIN JOHNSTONE, PhD
Professor

TERI KLEIN, PhD
Professor & Associate
Chair of Faculty Affairs

YING LU, PhD
Professor

AARON NEWMAN, PhD
Associate Professor

JULIA PALACIOS, PhD
Associate Professor

MANUEL RIVAS, DPhil
Assistant Professor

CHIARA SABATTI, PhD
Professor & Associate
Chair of Education and
Training

JULIA SALZMAN, PhD
Associate Professor

LU TIAN, ScD
Professor & Director of
Graduate Studies for Master of
Science Programs

ROBERT TIBSHIRANI, PhD
Professor

WING HUNG WONG, PhD
Professor

SERENA YEUNG-LEVY, PhD
Assistant Professor

JAMES ZOU, PhD
Associate Professor

NIMA AGHAEEPOUR, PhD
Associate Professor

RUSS B. ALTMAN, MD, PhD
Professor

EUAN A. ASHLEY, MRCP, DPhil
Professor and Associate Dean
of the School of Medicine

AKSHAY CHAUDHARI, PhD
Assistant Professor

JONATHAN CHEN, MD, PhD
Assistant Professor

CHRISTINA CURTIS, PhD
Professor

MANISHA DESAI, PhD
Professor

OLIVIER GEVAERT, PhD
Associate Professor

TINA HERNANDEZ-BOUSSARD, PhD
Professor

CURTIS LANGLOTZ, MD, PhD
Professor

STEPHEN B. MONTGOMERY, PhD
Professor & Director
of Graduate Admissions

MARK MUSEN, MD, PhD
Professor

NIGAM H. SHAH, MBBS, PhD
Professor

DENNIS WALL, PhD
Professor & Director
of Graduate Studies

JOHN WITTE, PhD
Professor

EMERITUS FACULTY | DEPARTMENT OF BIOMEDICAL DATA SCIENCE

DANIEL BLOCH, PhD

PHILIP W. LAVORI, PhD

DANIEL RUBIN, MD

ALICE S. WHITTEMORE, PhD

COURTESY FACULTY
DEPARTMENT OF BIOMEDICAL DATA SCIENCE

EHSAN ADELI, PhD
Assistant Professor

ANDREW GENTLES, PhD
Associate Professor

ZIHUAI HE, PhD
Assistant Professor

JOHN P.A. IOANNIDIS, MD, DSc
Professor

MIRABELA RUSU, PhD
Assistant Professor

ADJUNCT FACULTY
DEPARTMENT OF BIOMEDICAL DATA SCIENCE

CARLOS BUSTAMANTE, PhD
Adjunct Professor

FRANCISCO M. DE LA VEGA, DSc
Adjunct Professor

MOHIT KAUSHAL, MD
Adjunct Professor

MATT LUNGREN, MD
Adjunct Professor

VIJAY PANDURANGAN, MSc
Adjunct Professor

MEI-CHIUNG SHIH, PhD
Adjunct Professor

TANVEER SYEDA-MAHMOOD, PhD
Adjunct Professor

Department of Biomedical Data Science, 2024 DBDS Annual Retreat

Help Us Advance AI and Biomedical Data Science for a Healthy Future
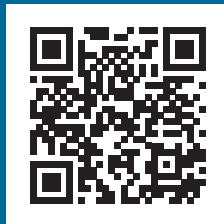
Advances in AI and biomedical data science are unlocking new possibilities for improving human health — but how we build and deploy these tools will shape whether the benefits are equitable and far-reaching. DBDS is redefining what a modern academic department can be: an engine of AI-driven discovery at Stanford.

Our research community is embedded at the heart of one of the world's top medical schools at one of the world's top universities — with unique access to clinical data, cutting-edge technology, and collaborators across academia, biopharma, healthcare, and technology. From multimodal foundation models to agentic AI systems, our faculty and students are creating tools that are creating the path to precision health and medicine.

DBDS is also growing the future — training bilingual scientists fluent in both computation and biomedicine. Your support can help us expand this impact through fellowships, computing resources, faculty endowments, and strategic partnerships.

We invite you to invest in the next generation of data-driven medicine. Scan the QR code to learn more about supporting our work. Together, we can build what will make a healthier future for everyone.

DBDS

E N D N O T E S

1. Johri S et al. An evaluation framework for clinical use of large language models in patient interaction tasks. Nat Med. 2025 Jan;31(1):77-86. PMID: 39747685.

2. Raji ID et al. It's Time to Bench the Medical Exam Benchmark. NEJM AI. 2025;2(2): doi:10.1056/AIe2401235.

3. Zack T et al. Assessing the potential of GPT-4 to perpetuate racial and gender biases in health care: a model evaluation study. Lancet Digit Health. 2024 Jan;6(1):e12-e22. doi: 10.1016/S2589-7500(23)00225-X. PMID: 38123252.

4. Van Veen D et al. Adapted large language models can outperform medical experts in clinical text summarization. Nat Med. 2024 Apr;30(4):1134-1142. doi: 10.1038/s41591-024-02855-5. PMID: 38413730.

5. Goh E et al. Large Language Model Influence on Diagnostic Reasoning: A Randomized Clinical Trial. JAMA Netw Open. 2024 Oct 1;7(10):e2440969. doi: 10.1001/jamanetworkopen.2024.40969. PMID: 39466245.

6. https://som-shahlab.github.io/ehrshot-website/

7. https://som-shahlab.github.io/inspect-website/

8. https://som-shahlab.github.io/medalign-website/

9. https://crfm.stanford.edu/helm/

10. https://crfm.stanford.edu/

11. https://hai.stanford.edu/news/holistic-evaluation-of-large-language-models-for-medical-applications

12. Somani S et al. Using large language models to assess public perceptions around glucagon-like peptide-1 receptor agonists on social media. Commun Med (Lond). 2024 Jul 10;4(1):137. doi: 10.1038/s43856-024-00566-z. PMID: 38987347.

13. Kang M et al. Mapping single-cell developmental potential in health and disease with interpretable deep learning. bioRxiv [Preprint]. 2024 Mar 21:2024.03.19.585637. doi: 10.1101/2024.03.19.585637. PMID: 38562882 (accepted in Nature Methods).

14. Sun ED et al. Spatial transcriptomic clocks reveal cell proximity effects in brain ageing. Nature. 2025 Feb;638(8049):160-171. doi: 10.1038/s41586-024-08334-8. PMID: 39695234.

15. https://clinicalgenome.org/

16. Bridges SL Jr et al. Curating Genetic Associations With Rheumatologic Autoimmune Diseases to Improve Patient Outcomes. Arthritis Rheumatol. 2024 Nov;76(11):1577-1581. doi: 10.1002/art.42943. PMID: 38965695.

17. Verma A et al. Cellular behavior analysis from live-cell imaging of TCR T cell-cancer cell interactions. bioRxiv [Preprint]. 2024 Nov 21:2024.11.19.624390. doi: 10.1101/2024.11.19.624390. PMID: 39605616.

18. Hoffmann TJ et al. Genome-wide association study of prostate-specific antigen levels in 392,522 men identifies new loci and improves prediction across ancestry groups. Nat Genet. 2025 Feb;57(2):334-344. doi: 10.1038/s41588-024-02068-z. PMID: 39930085.

19. Pizurica M et al. Digital profiling of gene expression from histology images with linearized attention. Nat Commun. 2024 Nov 14;15(1):9886. doi: 10.1038/s41467-024-54182-5. PMID: 39543087.

20. Bouchard G et al. A quantitative spatial cell-cell colocalizations framework enabling comparisons between in vitro assembloids and pathological specimens. Nat Commun. 2025 Feb 6;16(1):1392. doi: 10.1038/s41467-024-55129-6. PMID: 39915493.

21. van Eijk RPA et al. Cultivating Patient Preferences in ALS Clinical Trials: Reliability and Prognostic Value of the Patient-Ranked Order of Function. Neurology. 2024 Jul 23;103(2):e209502. doi: 10.1212/WNL.0000000000209502. PMID: 38875513.

22. Zhang E et al. LLM-Lasso: A robust framework for domain-informed feature selection and regularization. arXiv Preprint. 2025 Feb 15;arXiv:2502.10648.

23. Liu Q et al. An encoding generative modeling approach to dimension reduction and covariate adjustment in causal inference with observational studies. Proc Natl Acad Sci U S A. 2024 Jun 4;121(23):e2322376121. doi: 10.1073/pnas.2322376121. PMID: 38809705.

24. Gablenz P, Sabatti C. Catch me if you can: signal localization with knockoff e-values. J R Stat Soc Series B Stat Methodol. 2024 Jun 14;87(1):56-73. doi: 10.1093/jrsssb/qkae042. PMID: 39935679.

25. Cappello L et al. An efficient coalescent model for heterochronously sampled molecular data. J Am Stat Assoc. 2024;119(548):2437–2449. doi:10.1080/01621459.2024.2330732.

26. Cui Y et al. Inferences for the distribution of the duration of response in a comparative clinical study. Clin Trials. 2024 Oct;21(5):541-552. doi: 10.1177/17407745241264188. PMID: 39114952.

Stanford University School of Medicine
Department of Biomedical Data Science
Edwards Building
300 Pasteur Drive
Room R348
Stanford, CA 94305-5464

Department of Biomedical Data Science

STANFORD MEDICINE

dbds.stanford.edu