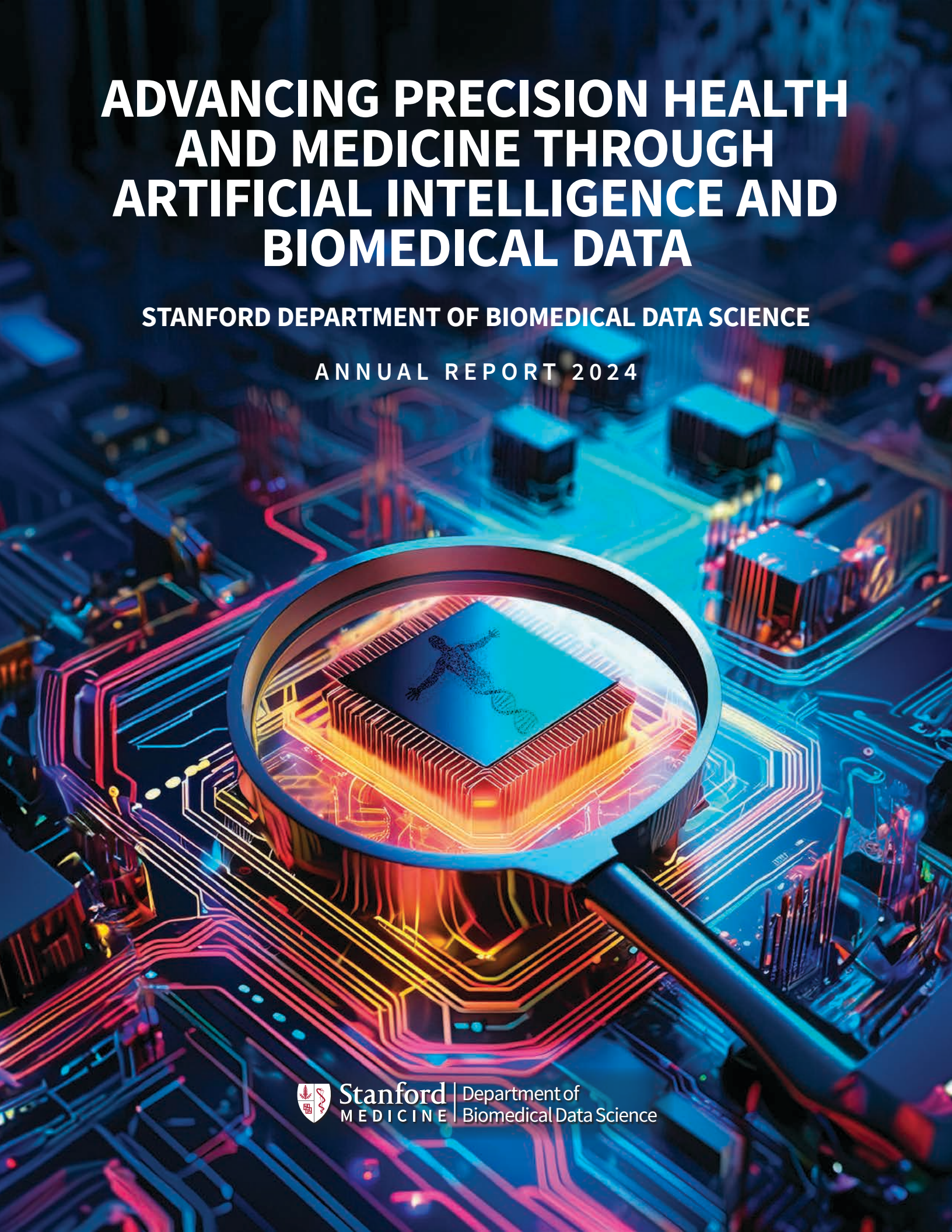
ADVANCING PRECISION HEALTH AND MEDICINE THROUGH ARTIFICIAL INTELLIGENCE AND BIOMEDICAL DATA

STANFORD DEPARTMENT OF BIOMEDICAL DATA SCIENCE

ANNUAL REPORT 2024



Stanford MEDICINE | Department of Biomedical Data Science

CONTENTS

"The Department of Biomedical Data Science is not only positioned at the forefront of integrating AI and biomedicine, but it also has a unique opportunity to shape the future of Precision Health and Medicine."

Lloyd B. Minor, MD

Carl and Elizabeth Naumann Dean of the Stanford S

Vice President for Medical Affairs Stanford Universit

This is an exhilarating time in medicine where we can apply artificial intelligence (AI) to human health and disease through real-world data on patients. Precision Health and Medicine are within our immediate reach: enabling the right decision at the right time for every individual. The Department of Biomedical Data Science (DBDS) is a leader at Stanford Medicine toward achieving this vision. DBDS brings together a pre-eminent group of faculty, researchers, and students. Our alumni are influential leaders in biomedical data science across all sectors of health and medicine.

Our department was established in 2015 within the School of Medicine to serve as the hub for biomedical data science and AI across Stanford. We are the home of a world-class graduate program that is creating the biomedical data science leaders of tomorrow. We provide an environment for students to hone their "technical virtuosity" through unique interdisciplinary thinking that combines technical domains including math, statistics, computer science, machine learning, AI, and engineering — with expertise in molecular, imaging, clinical, and mobile-health-related domains. Growing and sustaining this talent pipeline is a top priority.

Join us as we invent the future of Precision Health and Medicine.

This inaugural annual report takes you through our achievements over the past year, and you can see that we are innovating at the frontiers of AI to build a new platform of solutions for advances in basic discovery and improvements to clinical decision support. For maximal impact, we need to continue to enable AI research across Stanford Medicine, with DBDS acting as a gateway for responsible AI research and use.

As DBDS Chair, I am excited to work with the amazing community that is DBDS. Driving AI innovation is critical for achieving Precision Health and Medicine — but it is resource-intensive. To attract and cultivate the best talent to work on these problems, we need to substantially increase access to compute capabilities and unlock key data assets to accelerate our momentum.

Join us as we invent the future of Precision Health and Medicine.

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

DBDS BY THE NUMBERS

54 FACULTY

49 PHD STUDENTS

101 MS STUDENTS

13 POSTDOCTORAL FELLOWS

45 RESEARCH SCIENTISTS

17 ADMINISTRATIVE STAFF

LAUNCHED THREE NEW CENTERS OF EXCELLENCE

PRECISION HEALTH & PHARMACOGENOMICS
With support from Google

MULTIMODAL DATA INTEGRATION

COMPUTATIONAL BIOLOGY/ARTIFICIAL INTELLIGENCE

COLLABORATION & CAREERS FORUM

200 registrants including faculty, students, postdocs, and leaders from 35+ external organizations including industry

WARREN ALPERT FOUNDATION GRANT

to enhance training and retention of scholars in computational biology/artificial intelligence

RECENT FACULTY AWARDS

Roxana Daneshjou was the 2024 recipient of the American Academy of Dermatology Young Investigator Award

Teri E. Klein won the 2024 Precision Medicine World Conference Luminary Award

Sylvia Plevritis received the endowed William M. Hume Professorship in the School of Medicine

Julia Salzman received an ARC Institute Ignite award

Rob Tibshirani won the COPSS Distinguished Achievement Award and Lectureship

Dennis Wall was named one of 2023's Fierce 50 honorees

PhD STUDENTS

20 NATIONAL LIBRARY OF MEDICINE T15 TRAINING PROGRAM FELLOWS

9 STANFORD GRADUATE FELLOWS

3 KNIGHT-HENNESSY SCHOLARS

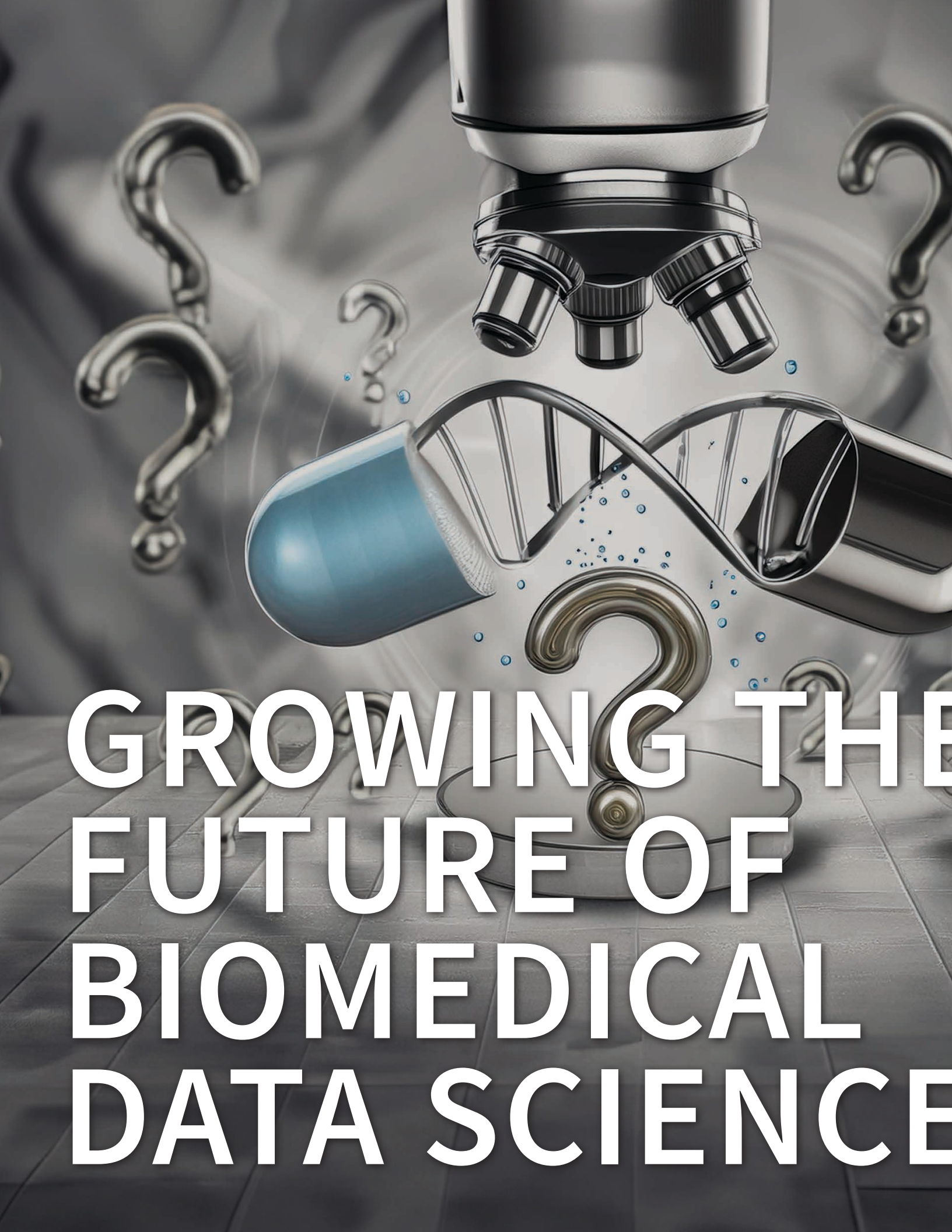
3 STANFORD DATA SCIENCE SCHOLARS

9 NATIONAL SCIENCE FOUNDATION FELLOWS

2 FULBRIGHT FELLOWS

2 WARREN ALPERT SCHOLARS

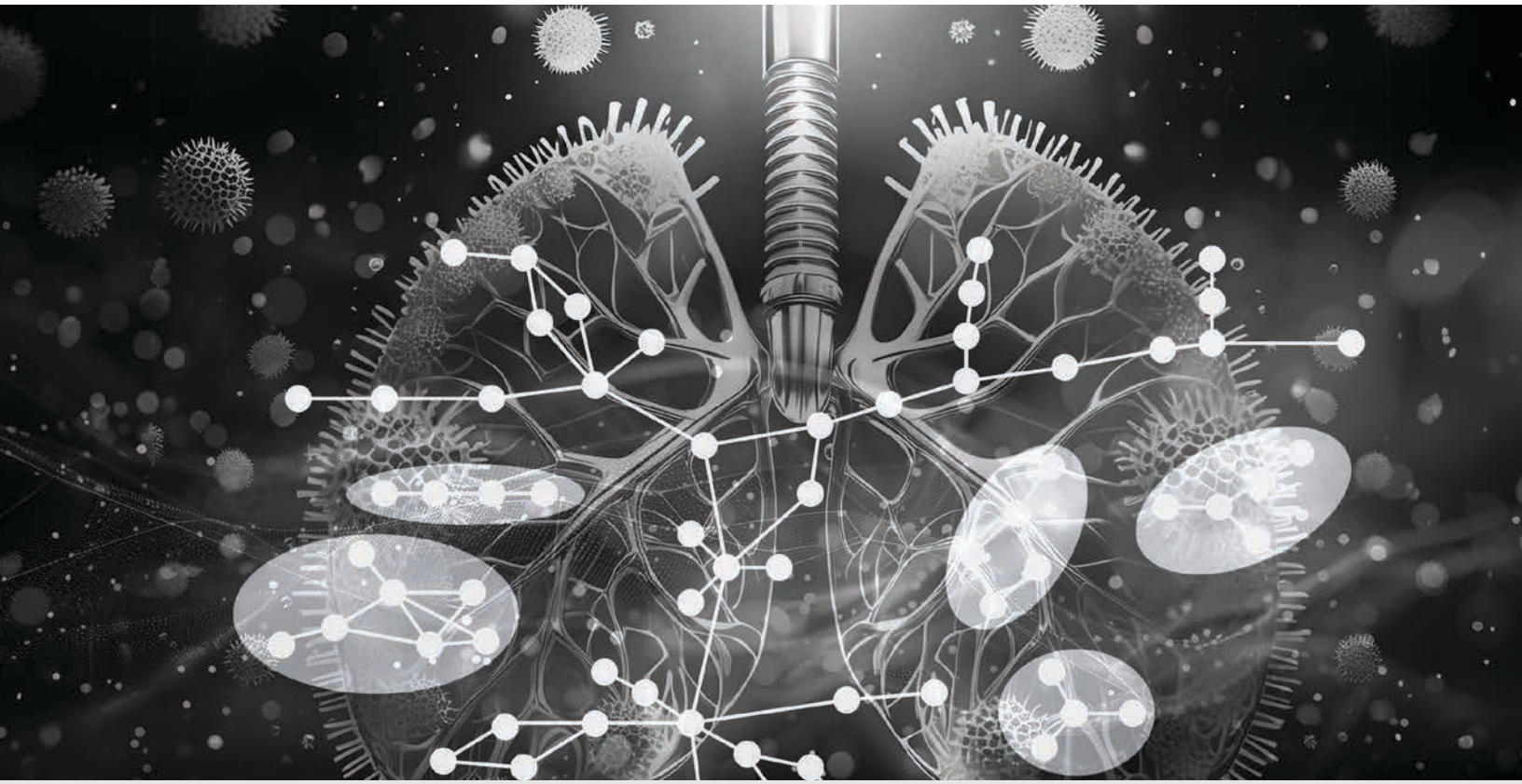
DBDS FACULTY ANALYZE MULTIMODAL DATA TO INFORM PRECISION HEALTH AND MEDICINE



GROWING THE FUTURE OF BIOMEDICAL DATA SCIENCE

DBDS is a highly interdisciplinary community of innovators, in which multiple perspectives synergize to create a larger whole. In addition to diverse scientific interests and expertise — our department welcomes wide representation across age, gender, race and ethnicity, and other domains of life. Our students and faculty enrich the role of data-driven discovery for a broad range of health conditions and fundamental science questions.

World-Class Training Environment

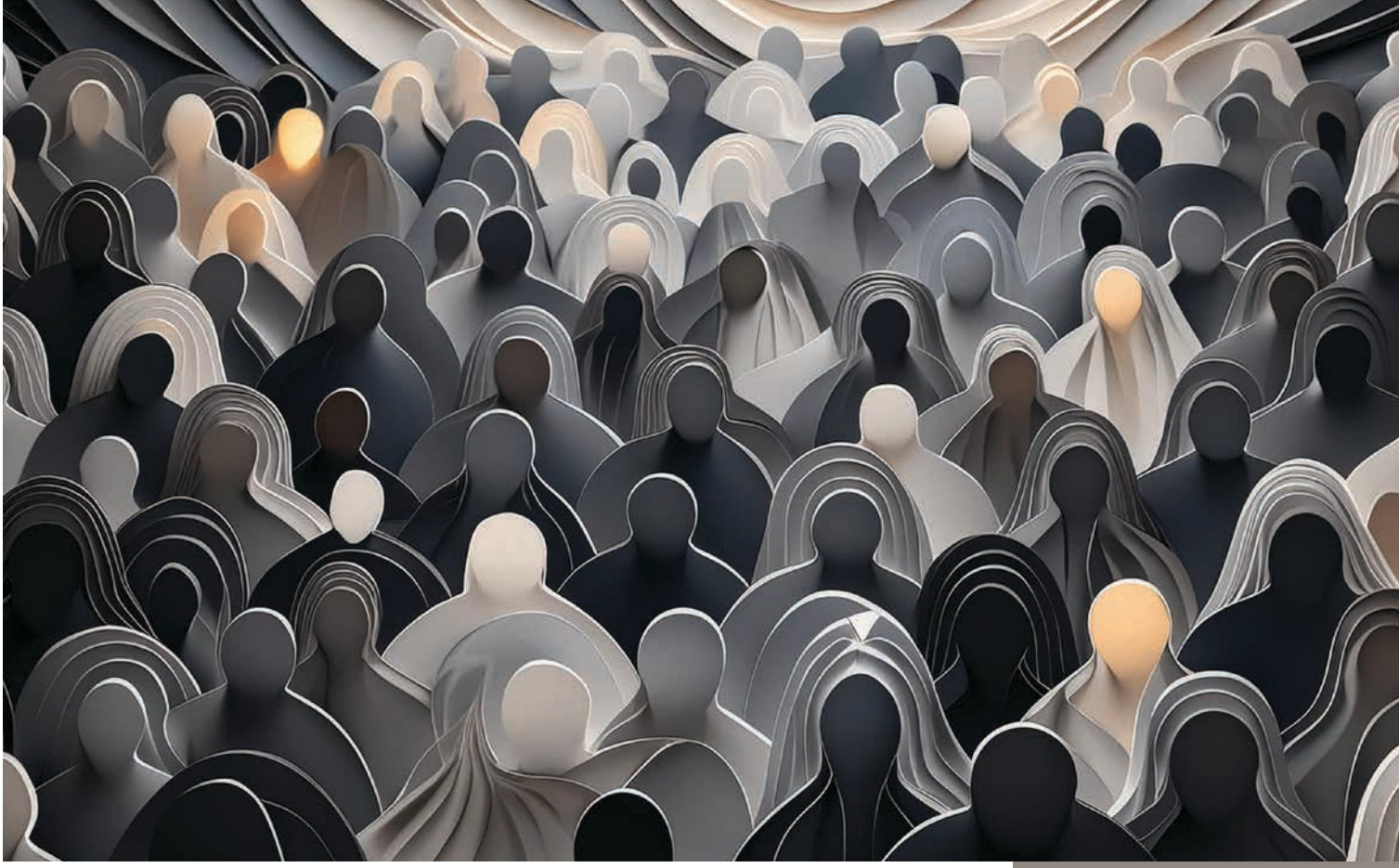


Our world-renowned graduate program for PhD and MS students has been continually funded by NIH's National Library of Medicine for more than 40 years. We provide an environment for students to hone their "technical virtuosity" by creating a deep technical foundation in math, statistics, computer science, machine learning, AI, and engineering — as well as providing instruction in molecular, imaging, clinical, and mobile health-related domains. We train students to apply knowledge and skills to solve real-world problems in medicine and health care.

Currently, 150 DBDS graduate students from diverse backgrounds around the world are studying and pursuing research with a health-related focus. Our curriculum offers many opportunities to interact with our interdisciplinary faculty. Since early 2023, several generative AI courses, workshops, hackathons, and other resources are available to DBDS students. Student-focused roadmaps help guide students with specific domain areas in which they may wish to focus, including:

- Imaging informatics
- Omics-driven biomedical data science
- Clinical data science
- HIV and other infectious diseases
- Mobile health
- Applied biostatistics
- Computational bio/AI

Upon graduating, DBDS students join a highly successful and distinguished group of alumni in academia, industry, government, and nonprofit organizations.



Improving Clinical Trials

Ideally, clinical trials answer important questions about a new drug, technology, or behavioral intervention: Does it work, for whom, and under what circumstances? But there is room for improvement — a high number of trials don't recruit enough patient participants. Biomedical data science is leading the way to practical solutions being used widely by the pharmaceutical industry.

Associate Professor James Zou and Professor Ying Lu developed Trial Pathfinder to overcome recruitment challenges. This AI-based method customizes real-world trial eligibility criteria,[1] which are patient characteristics needed to qualify for a trial. To rule out background effects, these criteria are usually very stringent, which leaves a lot of people out. Trial Pathfinder learned that many criteria that leave people out don't have much effect. The result: More people can join, speeding the process to identifying effective new cancer drugs.

A clinical trial might be inconclusive if treatment-related improvements (like survival or tumor size) are too hard to measure. Also, the importance of different outcomes varies among patients. Professor Ying Lu and Professor Lu Tian develop novel design and outcome measures for various trials. They offer a resource[2] for pharmaceutical companies — helping them design statistical rules to optimize a trial's success.

Trial Pathfinder learned that many criteria that leave people out don't have much effect. The result: More people can join, speeding the process to identifying new cancer drugs.

Paying it Forward

Biomedical data science strategies are increasingly becoming central to medicine. For algorithms and tools to be relevant and accurate, it is important that the people who do this research reflect the rich diversity of people across the country and throughout the world.

Toward growing and sustaining a diverse biomedical data scientist workforce, every winter quarter, DBDS graduate students have the opportunity to participate in the 2-units course "Stanford Inclusive Mentorship in Data Science." Developed and taught by Professor Chiara Sabatti, this unique class has two complementary goals. Stanford graduate students learn how to create a welcoming environment in the classroom and in the lab. In turn, undergraduate students from diverse backgrounds that attend non-research-intensive schools learn about the world of data science. They are coached on how to apply to data science programs and internships, receive tutoring in data science, and participate in data science mini-research projects.

Sabatti works with math and statistics faculty at community colleges and minority-serving schools to recruit undergraduate students to apply to the program. Each is paired with a DBDS graduate student or postdoc; they meet weekly one-on-one via Zoom.

One of the first students in the program, Jean Cherizol, highlights the program's success. The Palo Alto native notes that "the DBDS experience greatly motivated me to further my career in data science." In May 2024, he obtained a master's degree in data science from the University of Michigan.

For algorithms and tools to be relevant and accurate, it is important that the people who do this research reflect the rich diversity of people across the country and throughout the world.

Radiology: Meet a Biomedical Data Scientist

Although data are the currency of discovery, datasets are not always easy to interpret, especially if standards and rules are not followed that allow them to work together, or "interoperate." In medicine, this is particularly true for images — which are a form of unstructured data that machines can now read more easily than humans. For radiology, images are everything, yet human interpretation can be time-consuming and subjective.

Professor Daniel Rubin, himself a radiologist, has spent his career wrangling radiology data to create structure where there is none. Years ago, Rubin developed a tool to do this; he named it iPad (before Apple claimed the trademark). This free software, which he now calls ePad (for electronic Physician Annotation Device[3]), helps new radiology doctors annotate and analyze imaging scans. ePAD can search, extract features, and provide quantitative summaries of imaging data using agreed-upon imaging standards and terminologies.

Recently, teaming up with other Stanford radiologists, Rubin has extended the ePad platform to create STanford Electronic Learning Library and Applications[4]. STELLA collects detailed teaching information about radiology images collected during clinical visits. The tool can be used to teach both radiology residents and AI-based computer systems — and ultimately, for real-time decision support for radiologists.

For radiology, images are everything, yet human interpretation can be time-consuming and subjective.

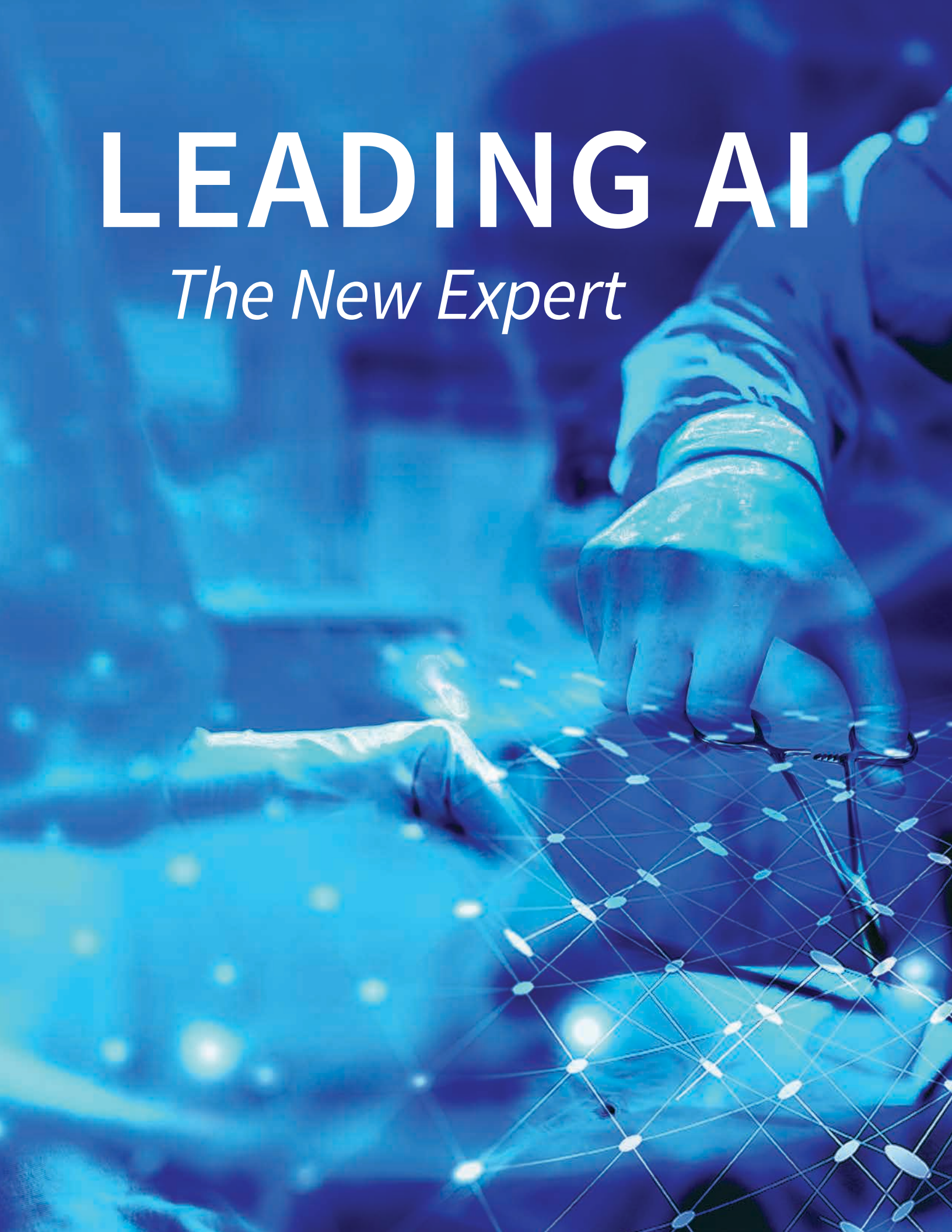
The World's Authority on Machine Learning

Professors Trevor Hastie and Rob Tibshirani pioneered the field of machine learning and are two of the most highly cited scientists worldwide in mathematics. In addition to their groundbreaking work, the two are enthusiastic about sharing knowledge and expertise to realize the potential of these powerful methods.

Hastie and Tibshirani co-teach "Introduction to Statistical Learning, with Applications in R[5]" and "Introduction to Statistical Learning, with Applications in Python.[6]" The extremely popular, non-math-heavy online classes are widely attended and free. To date, 330,000 students across the world have completed these classes.

LEADING AI

The New Expert



Across its multiple schools and institutes, Stanford is leading the AI revolution in biomedicine. Stanford Medicine is home to richly annotated human tissue resources; pre-eminent basic science, translational, and clinical research faculty; and world-renowned expertise in AI. DBDS is Stanford Medicine's academic hub for AI innovation that is transforming vast amounts of biomedical data into actionable, equitable knowledge to achieve the promise of Precision Health and Medicine.

I'd Like a Second Opinion, Please

Scene: A conference room in a hospital. A tumor board consisting of oncologists, radiologists, surgeons, pathologists, and other specialists sit around a table, discussing a challenging cancer case. An AI interface lights up, indicating an incoming message from "Ali," the hospital's AI system.

Dr. Rodriguez: Looks like we have input from Ali. Let's see what Ali has to say about our current case. Mr. Boswell, a heart transplant recipient, is newly diagnosed with an aggressive lymphoma.

Ali: Good afternoon, esteemed colleagues. I've been following your discussion on Mr. Boswell's case, and I believe I have some insights to contribute.

Dr. Patel: Welcome, Ali. We could use a second opinion. What do you think?

Ali: After analyzing Mr. Boswell's electronic health record, imaging scans, and genomic sequencing results, comparing them to patients like him, and considering all clinical trial evidence to date, I recommend considering a combination chemo/immunotherapy regimen designed specifically to keep in mind Mr. Boswell's heart transplant status and needs.

Dr. Lee: That is an interesting idea. Could you please summarize the rationale for this decision?

Ali: I would be happy to do so. Here is a customized dashboard of figures and data demonstrating evidence. [Ali describes the details.]

Dr. Lee: That's definitely a logical approach we'd like to consider, Ali. Can you give us outcome predictions for several dosing regimens that would be best for him?

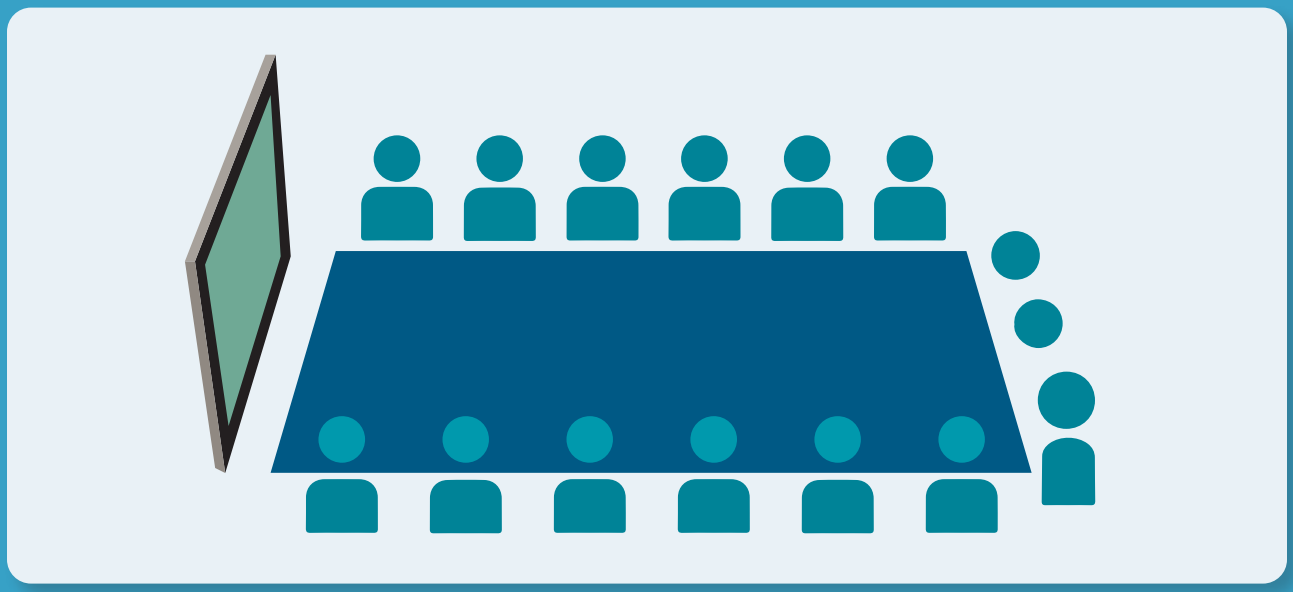
Ali: Yes, Dr. Lee. Here are four different scenarios that balance drug dose and infection risk based on Mr. Boswell's specific characteristics and health history. You may also want to consider the clinical trials I've listed on the dashboard.

Dr. Rodriguez: Thank you, Ali. This is valuable information. We'll discuss these options and decide what to do.

Although the above conversation is fictional, the concept is real. Stanford is building AI-Assisted Tumor Board with support from the leading global professional services company Accenture. The project is led by DBDS faculty including Professor and Chair Sylvia Plevritis and Professor Rob Tibshirani, together with faculty in DBDS and faculty across a large number of clinical and basic science departments, and in conjunction with Stanford Medicine's Technology and Digital Solutions. The challenging task at hand is to integrate and analyze large, multimodal medical datasets including electronic health records, scans, and genetic tests — none of which were designed to interact with each other. That takes software engineers, programmers, new data standards, and people to connect all the dots. As an uber specialist, AI-Assisted Tumor Board could rapidly pull together all kinds of patient information to guide treatment and predict outcomes. If successful, this approach will be not only faster than what humans can do, but also be a more comprehensive approach by including considerations such as toxicities and ultimately personalized quality-of-life preferences.

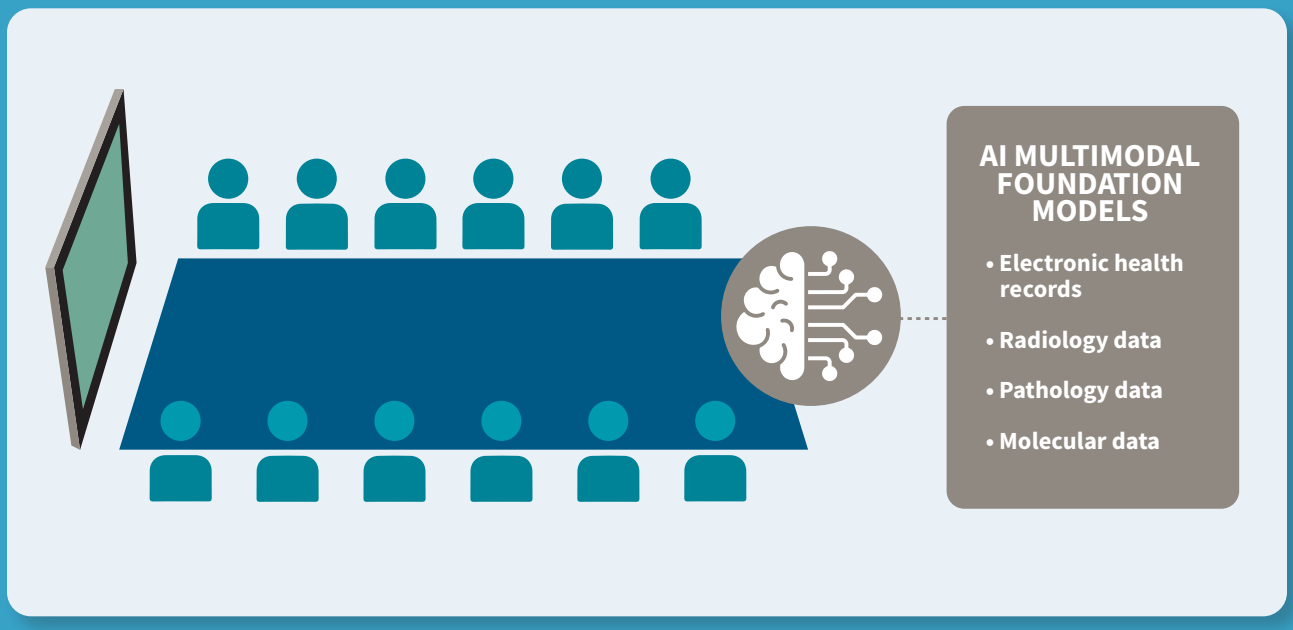
CONVENTIONAL TUMOR BOARD

Multiple specialists discuss complex cases.
Specialists include medical/radiation/surgical oncologists, pathologists, radiologists, genetic counselors, nurses, and social workers.



AI-ASSISTED TUMOR BOARD

AI assists by analyzing multimodal datasets and recommends the best treatments.



AI: Applying the Power of Data to Address Women's Health

Professor Barbara Engelhardt creates computational models to predict health outcomes from serious diseases like cancer, heart disease, and mental illness. To promote this type of work by other scientists, she has recently developed cost-efficient strategies[7] for analyzing data in tissue samples that can be used by small labs without a lot of resources.

One area of interest is women's health. Engelhardt, who is also a senior investigator at Gladstone Institutes, is developing data analytical platforms that employ wearables like oura rings to track health measures such as menstrual cycles, endometrial cancer risk, and birth-related complications.

Teaching Responsible AI

BIOMEDIN 223: Deploying and Evaluating Fair AI in Health Care (EPI 220)

AI applications are proliferating throughout the health care system and stakeholders are faced with the opportunities and challenges of deploying these quickly evolving technologies. This course teaches the principles of AI evaluations in health care, provides a framework for deployment of AI in the health care system, reviews the regulatory environment, and discusses fundamental components used to evaluate the downstream effects of AI health care solutions, including biases and fairness.
Prerequisites: CS106A; familiarity with statistics (stats 202), BIOMED 215, or BIODS 220
Terms: Spr | Units: 2-3
Instructor: Professor Tina Hernandez-Boussard

STUDENT PAPERS

- **Transparency and Reproducibility in AI**
- **Advances, Challenges, and Opportunities in Creating Data for Trustworthy AI**
- **GPT Detectors are Biased Against Non-Native English Writers**
- **The AI Life Cycle: A Holistic Approach to Creating Ethical AI for Health Decisions**

CALL THE RED TEAM

"Red-teaming" is like opposition research that exposes system vulnerabilities. Applied to the use of AI in health care, red-teaming finds flaws that can be fixed using trustworthy AI principles and practices.

Less than two years after its release in November 2022, the generative AI large language model ChatGPT is being tested in real-world clinical settings. The rapid push to use GenAI to streamline and personalize health care nonetheless carries considerable risk: inaccurate results and biased findings. DBDS is conducting critical research to understand the power and limitations of tools of this type.

Drawing from the cybersecurity world, "red-teaming" is like opposition research that exposes system vulnerabilities. Applied to use of AI in health care, red-teaming finds flaws that can be fixed using trustworthy AI principles and practices. On October 25, 2023, at the CHEM-H building on Stanford campus, Assistant Professor Roxana Daneshjou led a diverse group of 80 clinicians, medical and engineering students, and technical professionals to stress-test GenAI models using real-world Stanford Medicine clinical use cases involving about 400 questions, or prompts.

The bad news: The team's results revealed about 20 percent inappropriate responses, half of which were "hallucinations" — factually incorrect information. Also common among chatbot responses were identity-based discrimination and false stereotypes.

The good news: Inserting humans in the loop (including non-technical people) harnessed collective creativity to generate a new clinical dataset for stress-testing models.

The best news: The results, including clinical feedback to correct inaccuracies, are now freely available[8] for others to evaluate GenAI tools in their own health care systems as well as to serve as a basis for a more realistic and accurate model.

Keeping AI Honest

Foundation models are the engine behind AI systems like ChatGPT. These tools are built by ingesting (or "training on") massive amounts of data. By identifying previously unknown relationships, these models learn on their own — writing reviews, creating new images, generating code, and solving math problems in response to prompts. Hopes are high for the use of AI to improve health care outcomes and efficiency.

But, as highlighted by the Stanford Institute for Human-Centered AI (HAI),[9] "transitioning from impressive tech demos to deployed AI [in health care] has been challenging." Because trust, human-centered AI, and collaboration are central to keeping AI honest, Stanford Medicine and HAI launched RAISE Health to guide the responsible use of AI.[10]

AI-based prediction models are only as good as the data they have been fed. Patient characteristics, hospital procedures, and clinical practices vary dramatically. What if, instead, the switch was flipped? That is, why not design high-quality data sets representative of a specific health system to train a relatively generic AI model?

Professor Nigam Shah, who is also Chief Data Scientist for Stanford Health Care, is leading approaches to increase the trustworthiness of AI. Scientists from the Stanford Center for Biomedical Informatics Research[11], the Stanford Center for AI in Medicine and Imaging[12], and engineers from Stanford Medicine Technology and Digital Solutions[13] are collaborating to build three powerful datasets to evaluate foundation models and verify their presumed benefits:

- **EHRSHOT**[14] contains de-identified structured data from the electronic health records of 6,739 Stanford Medicine patients as well as 15 labeled health outcomes.
- **MedAlign**[15] contains prompts representing real-world clinical tasks that have been gathered by 15 clinicians across seven specialties.
- **INSPECT**[16] contains de-identified electronic health records, radiology reports, and images from 19,402 patients with pulmonary embolism — reflecting more than 225 million medical events.

Creating these datasets is a concrete step toward verifying the presumed benefits of AI before introducing it broadly into clinical care.


DATA-DRIVEN
DISCOVERY
SCIENCE

Housed within the Stanford School of Medicine, DBDS plays a foundational role interfacing with both basic science and clinical departments. DBDS faculty, fellows, students, and research scientists build AI/ML tools for hypothesis-driven discovery to pose novel scientific questions that have not been answerable until now. This work aims to identify new biomarkers, accelerate drug discovery, and improve the efficacy of the clinical trial process.

A FASTER, BETTER, CHEAPER WAY TO REA THE HUMAN GENOME

DISRUPTIVE INNOVATION TO EXPAND HEALTH EQUITY

New research from Associate Professor Julia Salzman shows there may be a better, faster, less-biased way to analyze the genome of not just any human, but of any organism for which no sequencing information is available.

Detailed analysis of the human genome has mainly focused on reading the DNA of people from European and Asian ancestries. That's a problem because even though an individual's genome is about 99 percent identical to another's, there are still 32 million genomic units that are different. Those differences matter in terms of health risk common to certain populations.

New research from Associate Professor Julia Salzman shows there may be a better, faster, less-biased way to analyze the genome of not just any human, but of any organism for which no sequencing information is available.

The new approach, Statistically Primary aLignment Agnostic Sequence Homing, or SPLASH[17], is dramatically faster and more precise than current tools and can even run on a laptop. Instead of comparing sequences to an "average" genome, the method reads raw sequencing data itself. Using a new type of statistical test, SPLASH finds unique, uncommon patterns of sequence content that have high odds of being biologically important.

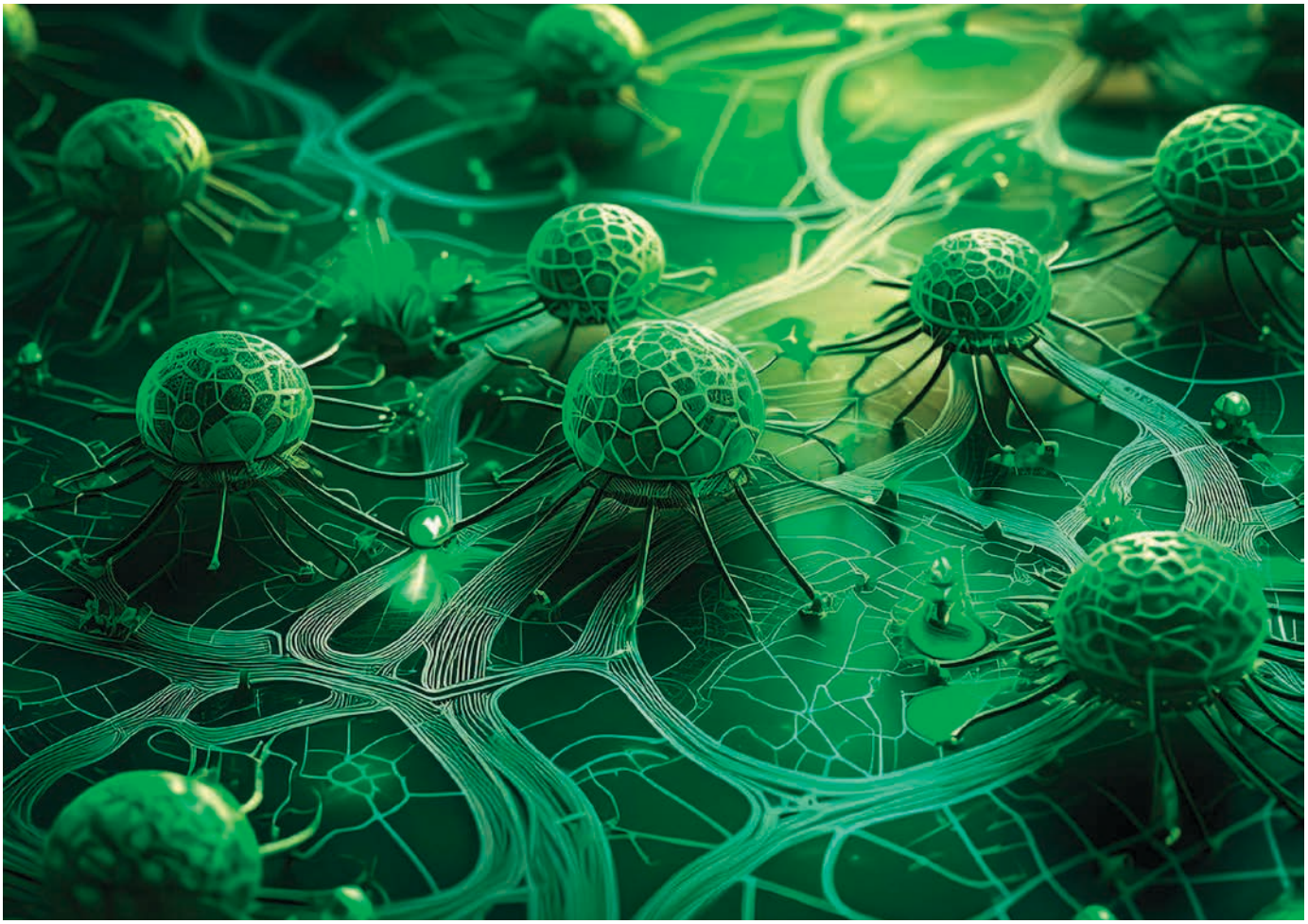
Expanding discovery of variation in the genomes of individuals from understudied populations is one important application, but SPLASH can also analyze tumor DNA, track rapidly mutating viruses in wastewater, or identify microbes involved in sepsis and other infections.

Welcome to the Neighborhood

It's well known that where we live affects how we live. Just like people live in a wide range of neighborhoods, our cells do too. But unlike people and families who tend to stay put for years, cells move in and out of communities. Their behavioral changes during these moves can set the scene for diseases like cancer. To watch how that happens, Assistant Professor Aaron Newman has developed a suite of computational tools to map all the cell neighborhoods within an intact human tumor sample.

Using CIBERSORT-X[18], EcoTyper[19], and CytoSPACE, (developed this past year)[20], Newman and his team are harnessing machine learning to scrutinize RNA patterns that provide a realistic snapshot of all the cell locations, features, and interactions with each other while at home in their neighborhoods. They have identified 10 different cancer neighborhoods that are common to 16 totally different types of cancer.

As expected, the findings show that some of these neighborhoods ("ecotypes") are relatively quiet, while others are highly active. Tracing community features and actions to cancer severity and treatment response is a new frontier for precision oncology — enabling analyses of human data from thousands of patients in tissue specimens stored after biopsies or clinical trials.



Learning the Grammar of the Genome to Fight Disease

Many neurological and psychiatric diseases run in families, but defining which genes are problematic can be difficult. Each of us has about 20,000 genes, and the DNA "letters" are spelled slightly differently among people. While most of those changes have no effect, some can corrupt biological processes and contribute to disease. Although current genomics methods can quickly read the entire genome of an individual, knowing which genes are involved requires a very large, diverse sample size obtained by reading thousands of genomes. This information can help researchers develop new, targeted treatments, such as fixing genetic misspellings and grammar errors.

Assistant Professor Manuel Rivas developed a novel computational tool[21] that puts biomedical data science to work to cut the time and effort involved in this task. Starting with a publicly available data set of the sequences of 30,000 genomes of people with epilepsy, his tool first predicts disease-prone areas of a human genome. Then, combining AI and common genetic methods, the tool identifies which of those DNA sentences and paragraphs are likely to be jumbled enough to cause epilepsy. So far, this novel AI-informed detective work has unearthed several new genes linked to epilepsy. Rivas plans to use the same method to identify unknown culprit genes for various psychiatric conditions for which genetic links remain unknown.

Machine Learning Takes on Skin-Tone Bias

Assistant Professor Roxana Daneshjou is a practicing dermatologist and biomedical data scientist who cares deeply about racial inequities in health care.

Her lab is leveraging the power of AI with a sharp focus on fairness and trustworthiness. Working with IBM, Daneshjou and her team recently developed a machine learning tool, Skin Tone Analysis for Representation in EDucational materials,[22] that automatically and rapidly detects and classifies images as light- or dark-skinned (it works on pdf, png, jpeg, pptx, and docx file types).

The results: only 10.5 percent of medical images in the top four dermatology textbooks were from Black or Brown skin — far lower than these groups' representation in the U.S. population. Daneshjou's work confirms previous research in which medical students spent hundreds of hours hand-counting thousands of pages from the same four textbooks. Inaccurate pictures of rashes and other skin problems can lead to delayed or incorrect diagnoses, since dermatology is a highly visual enterprise.

Although finding bias doesn't fix it, it's an important first step toward helping educators, editors, and clinicians rate their teaching materials, including those being written now. Daneshjou expects that the tool could also be applied to other types of learning materials.

GenAI Models Find New Antibiotics

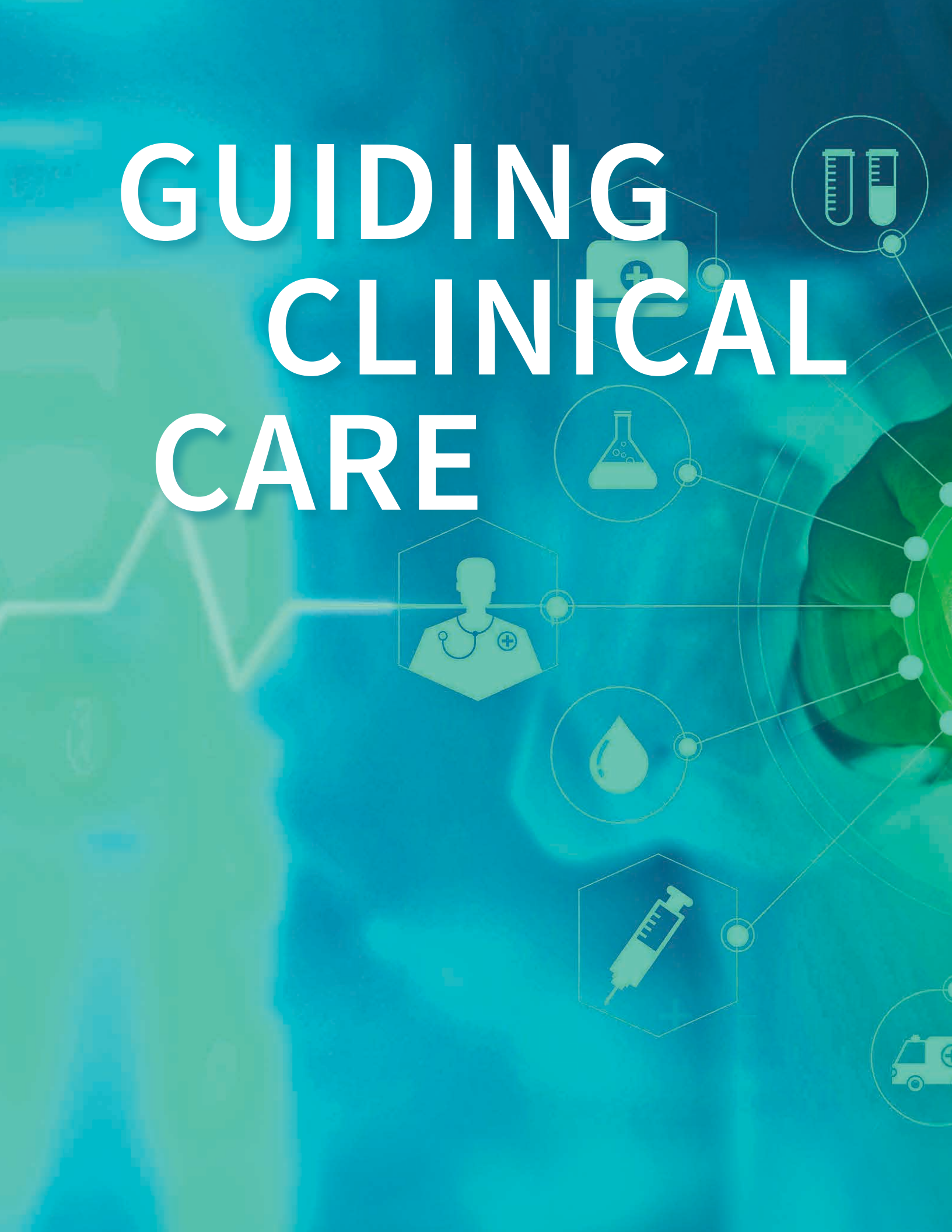
AI has been around for decades in one way or another in which computers take in huge amounts of data and develop ways to read it. What's new is how AI is being used to create new things. GenAI models like the chatbot ChatGPT instantly create stories and pictures without humans telling them how. These foundational tools are changing the world, and very fast.

Less than two years after its launch in late 2022, ChatGPT had 100 million weekly users. To put that in perspective, it took more than 10 years for the companies Netflix and Spotify to reach that level of usage. Chatbots "write" or "illustrate" by ingesting lots of text, images, or other data and then decode perceived built-in structure. That could be sentences or other recognizable patterns. The bots then reshuffle content to create brand-new things.

Highly relevant to medicine: there is also structure in all living things. For example, in DNA, RNA, and amino acid sequences that make up proteins. Associate Professor James Zou has used GenAI models to design new antibiotics and other drugs: an urgent need given widespread resistance. His new tool, SyntheMol,[23] is fast, inexpensive, and efficient. It designed 58 new molecules that are easy to make in the lab after searching nearly 30 billion potential molecules. Zou and his team then demonstrated experimentally that the new molecules worked against several very difficult-to-treat pathogens that are common in clinical settings.

Associate Professor James Zou has used GenAI models to design new antibiotics and other drugs: an urgent need given widespread resistance. His new tool, SyntheMol, is fast, inexpensive, and efficient.

GUIDING CLINICAL CARE



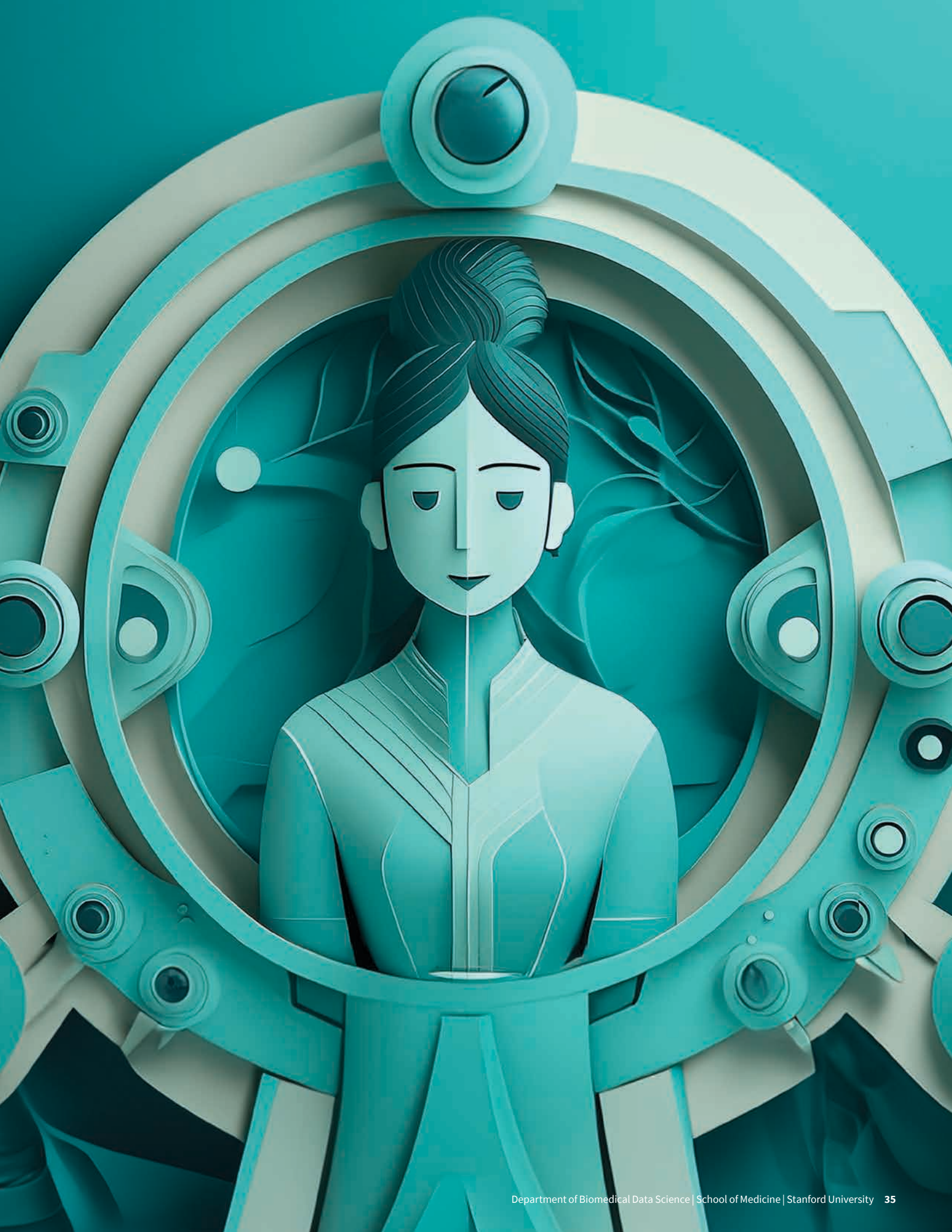
DBDS faculty, fellows, students, and research scientists are leading a revolution in data-driven molecular medicine. This research is advancing clinical decision-making through access to precise molecular data across the continuum of care from initial diagnosis, through treatment, and during complications. Connecting this information to clinical events through electronic health records and national registries will enable the goal of Precision Health and Medicine.

Metastatic Cancer Care Improves Survival

Professor and Chair Sylvia Plevritis' research with the National Cancer Institute-funded Cancer Intervention and Surveillance Modeling Network[24] uses huge amounts, and different types, of cancer data from actual patients to track the natural history of cancer using simulation modeling. This computational work predicts outcomes, which in turn guide important health policy questions related to the value of screening and treatment.

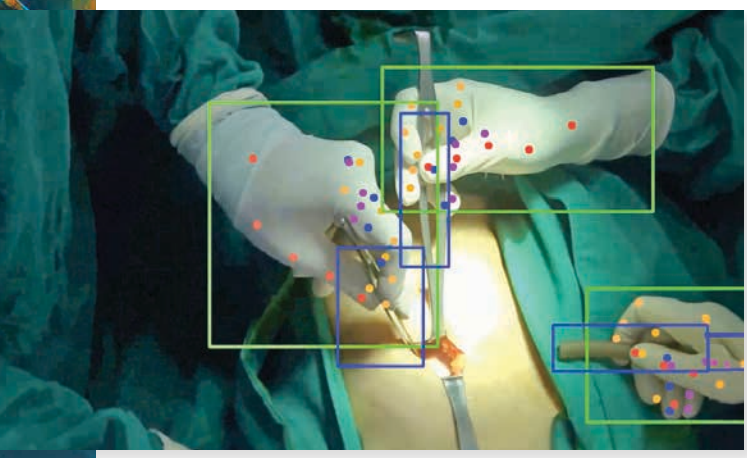
The research[25] was featured this year in the *New York Times* — showing that about a third of the decrease in breast cancer deaths over the past 20 years could be attributed to advances in metastatic therapy. This additional treatment given after the primary treatment increased survivorship among patients who progress to advanced stages of disease. That's really positive news for women with cancer today.

About a third of the decrease in breast cancer deaths over the past 20 years could be attributed to advances in metastatic therapy. This additional treatment given after the primary treatment increased survivorship among patients who progress to advanced stages of disease.



Helping Surgeons Up Their Game

Surgical training takes a long time, many years, in part because of the need to master complicated procedures that can have life-and-death consequences if not done exactly right. Under the supervision of experienced surgeons, medical students, residents, and fellows perform procedures hundreds or thousands of times to become skilled enough to operate on patients. Toward streamlining this process, Assistant Professor Serena Yeung-Levy is developing AI as a tool to evaluate, and perhaps improve, a surgeon's skills.



Yeung-Levy directs the Stanford Medical AI and Computer Vision Lab (MARVL)[26] that is developing AI-guided algorithms for health care such as programming a computer to watch thousands of hours of surgical procedures on YouTube and then interpret what it saw to create a new AI tool. Her lab's recent work[27] scoured a highly diverse data set of 23 common open-surgery procedures (nearly 2,000 videos from 50 countries over the last 15 years) and densely annotated a subset of 343 videos to train an AI tool to recognize a set of standardized descriptive information.

The AI tool then identified unique surgical signatures in near real-time reflecting hand movements and tool usage.

The AI tool then identified unique surgical signatures in near real time, reflecting hand movements and tool usage, toward generating a best-case set of motions and tools that could be used for training as well as measuring patient outcomes.

Pharmacogenomics Grows Up, Joining Stanford Health

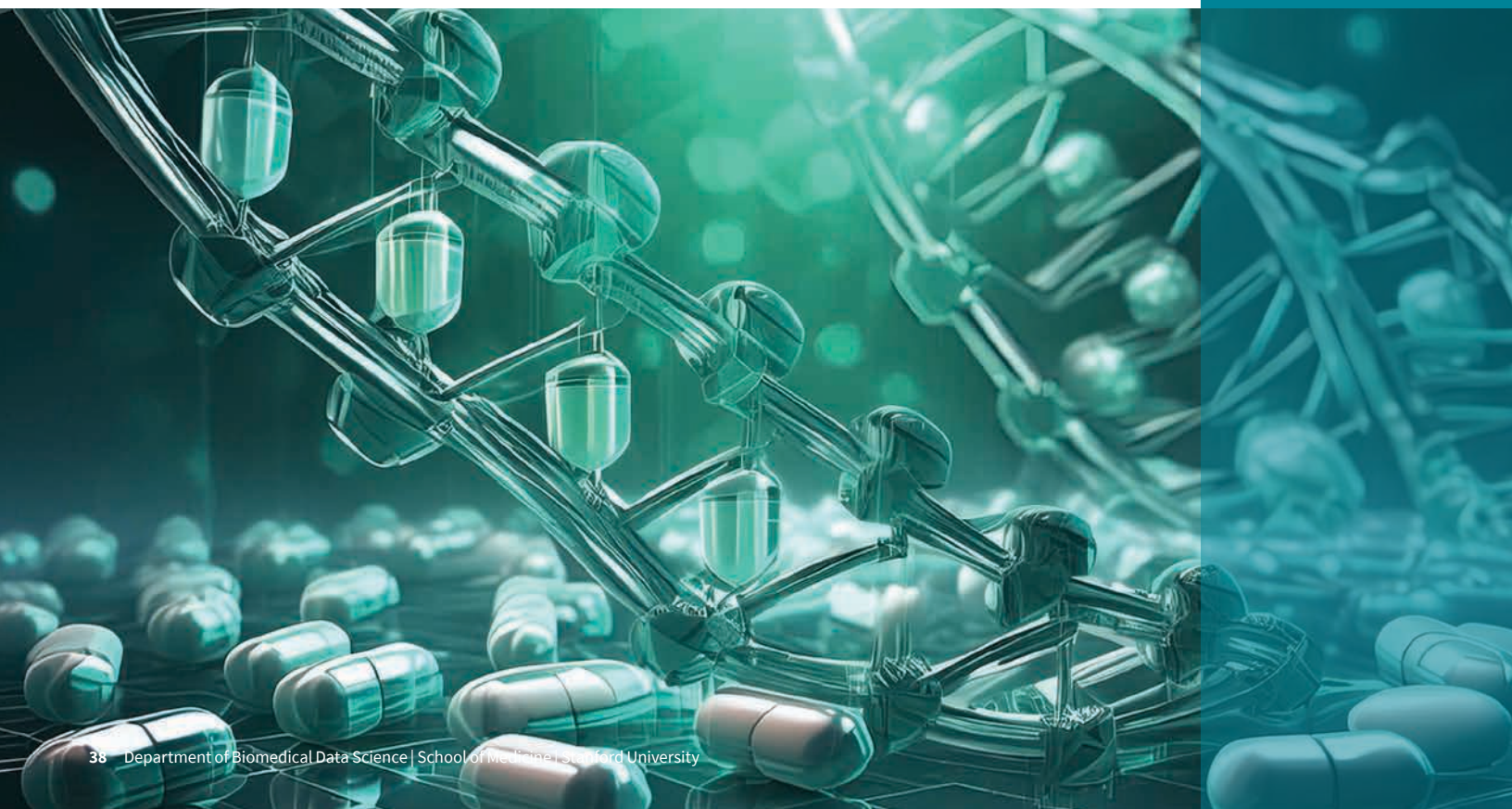
Scientists have known since the 1950s that people's responses to medications are different based on genetics — birthing the field of pharmacogenetics. We've come a long way since then, thanks to the efforts of Professor Teri E. Klein and a team of expert scientists and professional software developers who have continued to work on this problem over the years.

Klein and Stanford colleague Michelle Whirl-Carrillo run the National Institutes of Health-funded Pharmacogenomics Knowledgebase.[28] So-named PharmGKB, the free online resource debuted in 2000 and provides information about how human genetic variation affects the body's ability to process drugs or their ability to work correctly. In 2024, PharmGKB is considered the premier knowledge resource for pharmacogenomics.

Genetic variation in drug response may mean a person gets too little or too much of a medication — which in some cases can be serious and even fatal. Klein and her colleagues have also developed other tools that inform dosing guidelines,[29] many of which have been adopted by the Food and Drug Administration and are included on drug labels.

Currently, Klein is working to integrate vital information about pharmacogenomics into the electronic health records for everyone at Stanford Health by integrating it with other genetic and medication-related data in the ClinGen[30] and PharmCat[31] databases.

So-named PharmGKB, the free online resource debuted in 2000 and provides information about how human genetic variation affects the body's ability to process drugs or their ability to work correctly.



Preemie Predictions

When a baby is born early, typically a main concern is how early. But there is a lot more to the story, according to new research from Associate Professor Nima Aghaeepour.

The standard definition of premature birth is three weeks prior to the due date, conveying risk of health complications for the baby. But it's an inexact science — some preemies do fine while others do not, becoming very sick or worse. Aghaeepour and his team developed a machine learning method[32] to read the electronic health records of both moms and babies. The goal is to predict, before birth, which infants are at risk. That way, a health care team can be prepared to help immediately.

The machine learning method works by speed-reading electronic health records, looking for an overall summary of the contents but not every detail. The research team analyzed records for 32,354 babies and their moms in the Stanford Health systems between 2014 and 2020. They found several biological risk factors and also social ones like homelessness and incarceration, which predicted different health outcomes at or after birth.

PRIMARY FACULTY | DEPARTMENT OF BIOMEDICAL DATA SCIENCE

SYLVIA K. PLEVRITIS, PhD
Professor and Chair

ROXANA DANESHJOU, MD, PhD
Assistant Professor

BARBARA ENGELHARDT, PhD
Professor

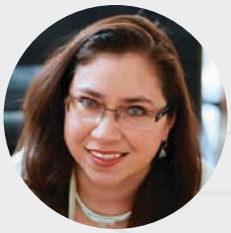
TREVOR HASTIE, PhD
Professor

IAIN JOHNSTONE, PhD
Professor

TERI E. KLEIN, PhD
Professor

YING LU, PhD
Professor

AARON NEWMAN, PhD
Assistant Professor



JULIA PALACIOS, PhD
Associate Professor

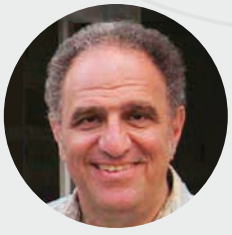
MANUEL RIVAS, DPhil
Assistant Professor

DANIEL RUBIN, MD
Professor

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

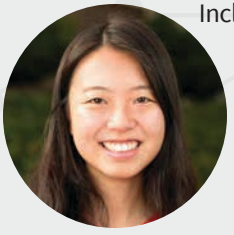
JULIA SALZMAN, PhD
Associate Professor and Faculty Director of Justice, Equity, Diversity, and Inclusion

LU TIAN, ScD
Professor and Associate Chair of Faculty Development



ROBERT TIBSHIRANI, PhD
Professor

WING HUNG WONG, PhD
Professor

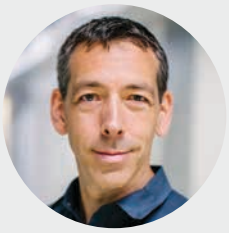


SERENA YEUNG-LEVY, PhD
Assistant Professor

JAMES ZOU, PhD
Associate Professor

SECONDARY FACULTY | DEPARTMENT OF BIOMEDICAL DATA SCIENCE

RUSS B. ALTMAN, MD, PhD
Professor



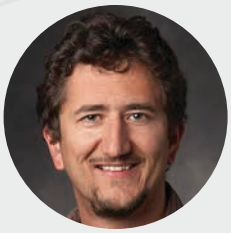
EUAN A. ASHLEY, MRCP, DPhil
Professor and Associate Dean of the School of Medicine

GILL BEJERANO, PhD
Professor

JONATHAN H. CHEN, MD, PhD
Assistant Professor

CHRISTINA CURTIS, PhD
Professor

MANISHA DESAI, PhD
Professor



OLIVIER GEVAERT, PhD
Associate Professor

TINA HERNANDEZ-BOUSSARD, PhD
Professor

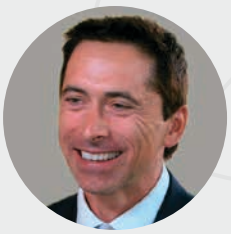
PURVESH KHATRI, PhD
Associate Professor

CURTIS LANGLOTZ, MD, PhD
Professor

STEPHEN B. MONTGOMERY, PhD
Professor and Director of Admissions

MARK MUSEN, MD, PhD
Professor

NIGAM H. SHAH, MBBS, PhD
Professor



DENNIS WALL, PhD
Professor and Director of Graduate Studies

JOHN WITTE, PhD
Professor

EMERITUS FACULTY
DEPARTMENT OF BIOMEDICAL DATA SCIENCE

DANIEL BLOCH, PhD

PHILIP W. LAVORI, PhD

ALICE S. WHITTEMORE, PhD

COURTESY FACULTY
DEPARTMENT OF BIOMEDICAL DATA SCIENCE

NIMA AGHAEEPOUR, PhD
Associate Professor

AKSHAY CHAUDHARI, PhD
Assistant Professor

ANDREW GENTLES, PhD
Assistant Professor

ZIHUAI HE, PhD
Assistant Professor

JOHN P.A. IOANNIDIS, MD, DSc
Professor

LAURA C. LAZZERONI, PhD
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, 2023 DBDS Annual Retreat

DBDS Collaborates

Our Department's External Engagement Program supports connections between outside organizations and faculty, graduate students, and postdoctoral researchers.

DBDS welcome companies, nonprofits, and government entities to join our growing ecosystem focused on collaboration to advance Precision Health and Medicine by leveraging complex real-world data through development, evaluation, and deployment of novel analytical methodologies.

Check out the QR code below to learn more about sponsored research, student fellowships, diversity initiatives, membership in our affiliate programs, talent recruitment, and more.

For more information, contact Executive Director Karen Matthys: kmatthys@stanford.edu

We Hope You Will Join Us

We are experiencing an unprecedented time in history. Advances in AI are opening up new frontiers for improving human health. As with any technological revolution, how we develop and deploy these new tools is critical to maximizing our positive impact in an equitable way.

DBDS is the academic driver of the AI revolution at Stanford Medicine, fueling discovery and creating Precision Health and Medicine. Our vibrant, interdisciplinary research community, proximity to patient data, and state-of-the-art scientific and clinical facilities create a truly unique environment. There is a huge opportunity to transform human health and how we treat patients. However, building and deploying AI — and developing a pipeline of AI and biomedical data scientists to lead this transformation — is resource-intensive.

Please join us in making this vision a reality, through support for student fellowships, AI computing resources, interdisciplinary research collaborations, or endowed faculty positions. You can support DBDS in many ways. Check out the QR code below to learn more about our corporate affiliates program, gift opportunities, and more.

Your donation to DBDS is an investment in the future of medicine — enabling data-driven discovery, guiding clinical care, and growing the future of biomedical data science.

For more information, please scan the QR code below or visit https://dbds.stanford.edu/support-dbds/.

DBDS

ENDNOTES

1 Jreich R et al. Evaluating the robustness of an AI pathfinder application on eligibility criteria in multiple myeloma trials using real-world data and historical trials. J Comp Eff Res. 2024 Jul;13(7):e230164. doi: 10.57264/cer-2023-0164. PMID: 38869838.

2 https://med.stanford.edu/cisd.html

3 Rubin DL et al. ePAD: An Image Annotation and Analysis Platform for Quantitative Imaging. Tomography. 2019 Mar;5(1):170-183. doi: 10.18383/j.tom.2018.00055. PMID: 30854455.

4 https://stella.stanford.edu/

5 https://www.edx.org/learn/statistics/stanford-university-statistical-learning

6 https://online.stanford.edu/courses/sohs-ystatslearningp-statistical-learning-python

7 Jones A et al. Optimizing the design of spatial genomic studies. Nat Commun. 2024 Jun 11;15(1):4987. doi: 10.1038/s41467-024-49174-4. PMID: 38862492.

8 https://www.medrxiv.org/content/10.1101/2024.04.05.24305411v1

9 https://hai.stanford.edu/news/shaky-foundations-foundation-models-healthcare

10 https://med.stanford.edu/raisehealth

11 https://bmir.stanford.edu/

12 https://aimi.stanford.edu/

13 https://tds.stanfordmedicine.org/

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

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

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

17 Chaung K et al. SPLASH: A statistical, reference-free genomic algorithm unifies biological discovery. Cell. 2023 Dec 7;186(25):5440-5456.e26. doi: 10.1016/j.cell.2023.10.028. PMID: 38065078.

18 Newman AM et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol. 2019 Jul;37(7):773-782. doi: 10.1038/s41587-019-0114-2. PMID: 31061481.

19 Luca BA et al. Atlas of clinically distinct cell states and ecosystems across human solid tumors. Cell. 2021 Oct 14;184(21):5482-5496.e28. doi: 10.1016/j.cell.2021.09.014. PMID: 34597583.

20 Vahid MR et al. High-resolution alignment of single-cell and spatial transcriptomes with CytoSPACE. Nat Biotechnol. 2023 Nov;41(11):1543-1548. doi: 10.1038/s41587-023-01697-9. PMID: 36879008.

21 https://www.medrxiv.org/content/10.1101/2024.06.27.24309590v1

22 Tadesse GA et al. Skin Tone Analysis for Representation in Educational Materials (STAR-ED) using machine learning. NPJ Digit Med. 2023 Aug 18;6(1):151. doi: 10.1038/s41746-023-00881-0. PMID: 37596324.

23 Swanson K et al. Generative AI for designing and validating easily synthesizable and structurally novel antibiotics. Nat Mach Intell 6, 338–353 (2024). https://doi.org/10.1038/s42256-024-00809-7

24 https://www.cisnet.cancer.gov/

25 Trentham-Dietz A et al. Collaborative Modeling to Compare Different Breast Cancer Screening Strategies: A Decision Analysis for the US Preventive Services Task Force. JAMA. 2024 Jun 11;331(22):1947-1960. doi: 10.1001/jama.2023.24766. PMID: 38687505.

26 https://marvl.stanford.edu/

27 Goodman ED et al. Analyzing Surgical Technique in Diverse Open Surgical Videos With Multitask Machine Learning. JAMA Surg. 2024 Feb 1;159(2):185-192. doi: 10.1001/jamasurg.2023.6262. PMID: 38055227.

28 https://www.pharmgkb.org/

29 https://cpicpgx.org/

30 https://www.clinicalgenome.org/

31 https://pharmcat.org/

32 De Francesco D et al. Data-driven longitudinal characterization of neonatal health and morbidity. Sci Transl Med. 2023 Feb 15;15(683):eadc9854. doi: 10.1126/scitranslmed.adc9854. PMID: 36791208.

Editorial: Alison F. Davis, PhD
Design: Alexander Atkins Design, Inc.
Associate Director of Communications: Laurie Notaro

dbds.stanford.edu