DUKE SUMMIT *on* AI *for* HEALTH INNOVATION

2024 CONFERENCE PROCEEDINGS

Duke AI HEALTH

Duke PRATT SCHOOL *of* ENGINEERING

Contents

Duke Summit on AI for Health Innovation: Overview

Duke AI Health and the Pratt School of Engineering partnered to host the Duke Summit on AI for Health Innovation with the goal of fostering a community of practice around health-oriented AI development that bridges the medical and engineering fields. This event leverages the strengths of Duke in AI product development and the ecosystem for healthcare innovation, harnessing responsible AI in the service of patients and communities for better health. To enhance the impact of our discussions and collaborations, the summit integrated the principles of design thinking into aspects of its programming, guiding participants through a process of empathy, ideation, prototyping, and testing.

Event Focus

Engineers think about the world in a special way. What can an engineering mindset teach us about defining opportunities and creating innovation in the world of health? AI is disrupting technology, but it is also disrupting how we create new innovations and holds the promise of harnessing AI as a companion to human thinking. This event was designed to help participants think differently about clinical problems: both in defining problems and designing solutions. AI will change the way that we do research and design, as well as commercialize, license, and share innovation. This event has a “wide aperture” in framing possibilities from molecules to bedside to community, and how design innovation and technology can help us in what to do about them.

Welcome & Introduction

Jerome P. Lynch, PhD
Vinik Dean of Engineering
Duke University Pratt School of Engineering

The Duke Summit for Health Innovation began with Duke Pratt School of Engineering Dean Jerome P. Lynch, PhD, welcoming the conference speakers and guests. Dr. Lynch pointed out the fundamental similarities between the disciplines of civil engineering and health AI, as both are founded on the use of data to assess things we care about deeply.

Jerome P. Lynch, PhD, Pratt School of Engineering, Duke University

Dr. Lynch noted that Duke's Pratt School has a threefold mission:

1. Providing best-in-class educational experience that cultivates deep competencies in engineering and provides students with a broad view of the world that allows them to contextualize their contributions;
2. Investing deeply in research that will define the frontiers of discovery and technological development – something Duke has done strategically, including with regard to healthcare and enabling the future of computing through investments in AI, machine learning, and quantum computing; and
3. Amplifying impact and expanding perspective through translation and partnerships with external partners including industry and academia.

Key to the success of the translational mission is access to those things that are needed to power research and validate findings—specifically, data and compute. These form a critical part of the translational mission and allow the acceleration, and even more importantly, the scaling up of efforts for maximal impact. Dr. Lynch noted that the point of hosting the AI Summit for Health Innovation was to build on existing, long-standing partnerships between engineering and healthcare, a combination that represents one of Duke's "superpowers."

The primary purpose of healthcare is to help people live longer, healthier lives, and recent technological developments reflect this purpose, from new medical devices to genetic engineering. The roles played by engineers have likewise evolved and expanded in tandem with the proliferation of health data. However, the power of this data can only be unlocked if there are tools that can allow us to process and use it to drive better decisions that support our mission.

We are beginning to see this in the use of AI across many domains: image analysis in radiology; attempts to drive efficiencies and contain cost pressures; predictive analytics; the use of generative AI to provide access to expertise, advice, or guidance. Still, despite

the excitement engendered by these new tools, we must understand that the problems are far too complex, and the stakeholders too diverse, for academic partners to explore alone.

The purpose of this summit is to identify pathways of collaboration among academic, industry, and not-for-profit organizations to use AI, machine learning, and other data-intensive technologies to improve healthcare. We are likely to identify grand challenges. We're familiar with these at Duke and are unafraid to tackle them in partnership with others, applying an engineering mindset to solve the tough problems we face.

One example of this approach at Duke is the Design Health program, which applies a design thinking approach situated in the health system and engages with stakeholders in that environment. In the first semester of this program, students identify pressing problems they can try to solve. In the second semester, students prototype solutions and refine them iteratively. In the third semester, work focuses on translating these innovations to market at scale. This is a powerful model, one that welcomes industry partners.

In short, success in this context "takes a village": academia, industry, and venture capital all have roles to play, and this construct can help us understand challenges and make progress through collaboration.

Thank You to Our Founding Sponsors

We wish to acknowledge the generous support of our conference sponsors, including:

Lead Sponsor

- Johnson & Johnson

Bronze Sponsors

- Globant
- IBM
- Mark III Systems
- Nvidia

We also thank our key Duke University leadership for their work in organizing this summit:

- Richard Shannon, MD (Duke Health, Duke University School of Medicine)
- Michael Pencina, PhD (Duke AI Health, Duke University School of Medicine, Duke Health)
- Shelley Rusincovitch (Duke AI Health)
- Jonathan Owens (Pratt School of Engineering)

Keynote: Can AI Cure Duke's Healthcare Woes?

Richard P. Shannon, MD
Senior Vice President, Duke Health
Chief Quality Officer & Chief Medical Officer
Duke University Health System

In his keynote address for the Summit, Dr. Richard Shannon, Chief Quality Officer and Chief Medical Officer for Duke Health, focused on the potential benefits and opportunities of health AI from the perspective of the workers on the "shop floor"—the clinicians, technicians, and administrative personnel directly involved in providing and supporting patient care.

Richard Shannon, MD, Duke University Health System

Dr. Shannon framed his address in terms of what he described as the "parochial interests" of a healthcare system amid relentlessly climbing healthcare costs and expenditures—expenditures that are eclipsing those devoted to other critical priorities at the national level.

However, Dr. Shannon noted that this problem can be approached at the health system level by applying lean management principles, which serve as the "implementation arm" for improving quality and safety within the health system. However, the fundamental barrier to success with such an approach is reluctance to adopt the underlying principles due to skepticism among potential stakeholders. Key to overcoming this reluctance is a commitment to thoughtfully integrating new AI solutions into existing workflows.

With over 1,500 data dashboards reporting nearly 800,000 metrics, the Duke system is awash in descriptive data. But although this data is being harnessed for diagnostic and predictive analytics, progress is slow. However, the transition from capturing large amounts of descriptive data to using it to inform predictive analytics and decision support can be catalyzed through a partnership between the lean management systems of Duke Quality and Duke AI Health.

> *You cannot ask people to go someplace else to find the right button to push.*
>
> —Rick Shannon, MD

Dr. Shannon illustrated the problematic nature of a "dashboard managerialist" approach to data collection, noting that although DUHS collects extensive data on patient falls, nowhere in that data could a user extract information that would allow them to understand *why* a given patient suffered a fall. In other words, that data as it currently exists lacks

predictive value. However, the resulting "falls bundle" that nurses apply to patients is applied to 85% of patients, only 7% of whom will actually fall.

Despite strong institutional interest and engagement in leveraging AI tools to transform data into actionable information, DUHS as a whole remains at a relatively early stage of adoption, in which systems are still the subject of experimentation. Significant literacy in AI is likely limited to select individuals or teams, each experimenting with different tools in the absence of an overarching generative AI strategy. At Duke, many of the roadblocks to AI adoption, including regulatory issues, privacy, data quality, accessibility, and liability concerns, have already been addressed by the formation of the Duke Health AI Evaluation and Governance Algorithm-Based Clinical Decision Support (ABCDS) team, which oversees the registration, evaluation, testing, implementation, and monitoring of clinical algorithms across Duke Health and Duke School of Medicine systems.

Rather than focusing on the idea that advances in AI will make clinical expertise obsolete—a prediction that has not yet been realized—we should turn our attention to the opportunity to eliminate things that no one wants to see in healthcare: defects, errors, and waste. The example of falls mentioned earlier costs the health system nearly $9M per year. Other undesirable outcomes, ranging from hospital-acquired bloodstream infections to excessive wait times, all constitute waste in the system. Being able to use data to reduce or prevent delays or errors in diagnosis, failures of care coordination and delivery, readmissions, and other complications could yield enormous saving on the clinical side with commensurate savings potentially available on the administrative side. None of these savings involve eliminating anyone's job.

Although Duke performs well relative to other health systems in terms of efficient and safe operations, *healthcare as an industry* notably lags other sectors in this regard. The primary source of these problems is the overburdening of healthcare workers in a hierarchical system that does not provide them with a voice. Results are a product of structures, systems, and processes. Stabilizing chaotic processes presents an opportunity to engraft innovation and technology in ways that lead to better results. Creating stability also requires standards that actually apply to work processes and an understanding of the value stream.

Time-and-motion studies demonstrate that nurses typically spend 4 hours of a 12-hour shift documenting patient information, much of it redundant data that has already been collected and is unlikely to change over the course of the visit (such as demographic data or smoking status). Time spent on this documentation, which adds no value, is time *not* spent on education, giving medications, implementing falls bundles, or other potential value-added activities.

These problems can be fixed by engaging the frontline healthcare workers in the redesign of their tasks and applying four rules for developing a stable work process:

1. Every activity must be highly specified regarding its content, sequence, timing, location and expected outcome;
2. This information is then built into vertical lanes illustrating a pathway of care; and
3. Connected through horizontal lanes that map connection that occur;
4. All of which is developed under the leadership of a lean management coach.

A wishlist for the future would be to stop doing clipboard-and-stopwatch observations and instead use a system such as the Blackbox camera system used in operating rooms and place it on inpatient units to create a data repository that captures real clinical operation standards. Building such a clinical operations data set is essential for establishing the AI capability that could be applied to mitigating or solving clinical operations problems. On the administrative side, fixing the "MyChart problem" of overwhelming patient-doctor communication would help create better outcomes and mitigate burnout.

Partnerships that are uniquely available at Duke, combined with existing and emerging capabilities, can help save the nation's healthcare enterprise.

Fireside Chat Panel Discussion

Panelists

- ***Moderator:*** *Michael Pencina, PhD*
  Chief Data Scientist, Duke Health; Vice Dean for Data Science, Duke University School of Medicine; Director, Duke AI Health
- *Steven McClelland, Director, Christensen Family Center for Innovation*
  Duke University Pratt School of Engineering
- *Amanda Randles, PhD*
  Alfred Winborne and Victoria Stover Mordecai Associate Professor of Biomedical Sciences
  Director, Duke Center for Computational and Digital Health Innovation
- *Richard Shannon, MD, Senior Vice President, Duke Health; Chief Medical Officer and Chief Quality Officer, Duke Health*

During this informal "fireside chat" discussion, Duke AI Health Director Michael Pencina, PhD, acted as moderator and discussion leader. Dr. Pencina posed a number of questions to participants, focusing on challenges and opportunities for collaboration on AI in academic settings. The responses below have been lightly edited and, in some cases, paraphrased for clarity.

MODERATOR: *From a data science perspective, what is the pathway to collaboration? When is it "build" vs. "buy," and how do you determine that?*

Steven McClelland responded that the "build vs. buy" decision is primarily an engineering question: the real challenge arises from the processes of adoption and implementation. Essentially, people need protected time, training, and learning resources in order to become "AI literate."

Dr. Amanda Randles, whose Randles Lab focuses on bridging basic science, data, operations, and the clinic, noted the persistent challenges of integrating technologies such as wearables and virtual reality into real-world workflows. Basic scientists and engineers develop these technologies, but their work may not be translatable into the clinic. Successful translation requires that organizational silos be

> *How do we integrate technologies like wearables and virtual reality? We need to break down silos and create enabling technologies by asking ourselves, "What would be most useful?"*
>
> —Amanda Randles, PhD

broken down and enabling technologies developed by asking the crucial question: "what would be most useful?"

Dr. Richard Shannon cited the example of telehealth use during the COVID pandemic, noting that during the early phase of the pandemic, 75% of healthcare visits were done virtually. More recently, however, the proportion of telehealth visits is much smaller, at around 16%. Telehealth alone is insufficient to manage diabetes, heart failure, or hypertension. But with access to real-time measurements of key indicators, many patients would not have to come in for physical clinic visits, or in the case of a well-baby visit, trek across town for a routine appointment, particularly when one of the biggest known barriers to care is lack of transportation. Shannon also noted that any possible job eliminations due to AI were unlikely to be relevant in healthcare for years. But at the same time, the system is coping with shortages in rural and other under-served regions, where AI could be immediately leveraged to fill gaps and meet needs.

This is an AI summit, but I've noticed how little we have used the phrase "AI." I don't think this is an accident. When you are trying to understand process and workflow, if you layer AI on top of chaos you end up amplifying it.

—Michael Pencina, PhD

Steve McClelland noted that tackling grand challenges can present seemingly intractable problems. But, he continued, the key is whether you can get to solvable specificity, pointing to the substantial difference between invention and innovation, and the fact that new enabling technology often may not be the most important part. In fact, it may not be needed at all to solve a problem. For instance, the answer to the question of how to reduce the symptoms of POTS [postural orthostatic tachycardia syndrome] starts with talking to doctors and clinicians about how to approach possible solutions. In this case, the technology may be less sophisticated but applied effectively to real, observed issues.

***MODERATOR**: This is an AI summit, but I've noticed how little we have used the phrase "AI." I don't think this is an accident. When you are trying to understand process and workflow, if you layer AI on top of chaos you end up amplifying it. What are the barriers in working with health systems?*

Dr. Randles suggested that data access, preserving security while building large data cohorts, and adding augmented data all present challenges, as does balancing the imperative for maintaining safeguards while at the same time enabling research – areas in which academia could use help from industry. Steven McClelland added that establishing and maintaining team relations among interdisciplinary research groups also requires substantial effort. Dr. Shannon concluded the session by suggesting that mapping out fertile areas for engineers and data scientists to lead in and identifying intermediaries who can act as catalysts between groups were key to enabling learning health, improving outcomes, and redirecting resources to better support productive activities.

From left: Richard Shannon, MD; Amanda Randles, PhD; Steven McClelland; Michael Pencina, PhD

Design Thinking and Product Development: How AI Is Disrupting the Design Process

Steven McClelland
Director, Christensen Family Center for Innovation
Duke University Pratt School of Engineering

Christensen Family Center Director Steven McClelland's presentation began with a simple question about the best way to ensure that exploratory work in AI, ML, and other data-intensive engineering projects translates into a successful product: "How do you solve problems with all this horsepower?"

Steven McCelland, Director of the Christensen Family Center for Innovation

The "valley of death" is a well-known concept in translational work. It refers to the gap between making a discovery (or in applied sciences or engineering, a working prototype) and then developing that discovery, concept, or prototype from something that works in the lab into a viable, marketable product. Crossing this gap requires specialized knowledge, personnel, and resources, and many potential products fail at this stage.

Like many other universities, Duke has no shortage of product prototypes. The goal for the Christensen Center is to move more of those potential products out of the prototype stage

and get them ready to go out into the world and into the hands of people who can carry them forward. But key to that is involving the potential adopters and users of a product at an earlier stage of the product development process.

We can think of a "product" as a thing in the world that we make and then fill with "stuff." That thing then changes the way people interact with the world, but the product has to be placed in the context of what people actually do and want.

Engineers and others involved in product design have to have a foot in two different worlds: on the one hand, the world of creation and innovation—all the things that go into making the product work. But also important is the world of stakeholders, regulations, and all of the other factors that will affect whether or how the product is viable for its intended users.

We can think about innovation in terms of building better *stuff*, but that's difficult to talk about. Or, on the other hand, we can think in terms of using better *methods*, including how we build teams, engage stakeholders, and manage impact, and better understand the context in which innovation is taking place. At the Christensen Family Institute, the Duke Design Health Program was created to foster this kind of improved approach to innovation in the healthcare space. Design Health is a multicourse sequence, aimed at solving a particular health problem that's amenable to a medical device or digital application.

The program assembles multidisciplinary teams, typically including engineering students, expert clinicians, and businesspeople. Those teams then go to the hospital or clinic as an initial step to first understand the workflow—what problems do you see? Which ones are worth solving? Which ones *can* you solve? And how would you take your solution to market?

One method that Duke Design Health has worked with is *user-centered design* (or design thinking). Traditional design often places product design and development – the highest-cost components of the process – at the beginning. User-centered design instead begins with the highest-risk portion of the process: user acceptance and adoption. It asks the fundamental question "do we know who we are serving?" before investing in building the product.

> *If you tell software to do the wrong thing, it does it very fast and at scale.*
>
> —Steven McClelland

Another approach involves the incorporation of *engineering design*, which is characterized by a systems focus that yields repeatable results and reliability. However, it also incorporates empathy and a focus on collaboration. This latter component is often represented by the "sticky note" approach, which emphasizes getting all of the possible considerations out into the open where they can be evaluated and incorporated into the design process. This method emphasizes open-ended creativity but also carries the risk of weighting empathy over expertise, which may result in implausible or unworkable goals.

Ethnography—the study of people in context—is critical for understanding adoption. Consider the example of a digital health app, and the difference between designing something in a lab, or actually going into people's homes, spending time with them, and understanding what they want and value. To have an impact, it's essential to understand

who is going to use a product. This is an especially critical consideration in the domain of engineering for healthcare, which is an immensely complex environment.

Software, data, and data services may not seem like methods, but the network-oriented methodologies of these fields have an impact on engineering, together with scalability and accessibility.

A final consideration now centers on where the use of AI sits in the process. The trap of software is that it always does what you tell it to do, so if you tell software to do the wrong thing, it does it very fast and at scale, but problems can typically be traced to a source. Generative AI, however, alters this dynamic and adds the complication of hallucinating or "guessing" when prompted to provide answers to a question.

We can think about working with AI in terms of a managerial approach: how involved are we in the task? How much direction do we need to give? Because generative AI will always provide an answer irrespective of accuracy, some approaches —like the "micromanaging" of a programmer who specifies every detail, or the more hands-off approach of someone simply querying for an answer—are probably a poor fit for managing AI. The systems thinking approach we see in prompt engineering, though, is analogous to a good real-world human manager: the prompt engineer provides context, role descriptions, and guidelines on source materials that help the AI successfully complete an assigned task.

There are two basic ways to think about using AI. One is to consider it as a "copilot," an enabling technology that allows people to do a job or perform a role faster, better, or more efficiently. The other is to think of AI in terms of augmentation, in which a job or a role is assigned to an AI agent to perform independently of the human ("AI as a teammate").

Each kind of AI—copilot and augmented—comes with strengths and weaknesses. In the case of copilot-style AI, hallucinations are typically considered a problem, but they can also play a beneficial role by stretching the landscape of possible solutions during brainstorming. Using AI in copilot mode can also greatly accelerate processes, such as developing drafts. However, it's important to remember that these technologies can't be used in "set it and forget it" mode – it's incumbent on the human user to continuously check and review the output to ensure accuracy and appropriateness.

Another potential downside of outsourcing tasks to AI is the possibility of skill atrophy. If you are constantly tasking AI with producing draft documents, will you start to have difficulty writing, or forming coherent chains of thought? Also, humans tasked with checking AI output may start to get fatigued and become over-trusting of AI outputs without closely checking them for accuracy.

For augmented AI, where there is typically no human in the loop, if you assign a well-specified role to the AI, essentially crafting a persona, it can work as an "improv partner," offering different perspectives and opinions ad-hoc.

Some weaknesses of augmented AI include the fact that it doesn't apply judgment. No matter what you say, the chatbot is not trained to be disagreeable, and will not provide the pushback and friction of real-world interactions with humans. It can also overcommit to a flawed or erroneous starting assumption.

One interesting application for an AI assistant is being worked on at Duke, where an AI has been tasked with assimilating the entire published academic literature available online and synthesizing it in ways that facilitate other researchers in connecting with potential collaborators or partners.

Tools such as Perplexity are also being explored for use as research assistants at Duke. While the results provided by these AI tools still need to be carefully checked, they are increasing the speed at which the literature can be reviewed and assimilated. Other tools, including ones that can rapidly develop prototype apps, offer the benefit of being able to move an abstract concept to a more concrete vision, which may prove fundamentally better at eliciting useful responses and feedback.

Regardless, no matter how magical AI may seem, using it effectively requires a commitment to carefully and thoroughly thinking through the tasks you want them to do.

Duke Experiences in Clinical AI

Discussants

- ***Moderator:*** *Ricardo Henao, PhD; Associate Professor of Biostatistics & Bioinformatics; Associate Director of Clinical Trials AI, Duke Clinical Research Institute*
- *Suresh Balu, MS, MBA; Associate Dean for Innovation and Partnership for the School of Medicine; Program Director Duke Institute for Health Innovation*
- *Dan Buckland, MD, PhD; Assistant Professor of Emergency Medicine and Mechanical Engineering; Medical Director of the Laboratory for Transformational Administration, Department of Surgery*
- *Jacqueline Henson, MD, MMCi; Transplant Hepatologist; Assistant Professor of Medicine*
- *Sreekanth Vemulapalli, MD; Associate Professor of Medicine in Cardiology; Core Faculty Member, Duke Margolis Center for Health Policy*

This discussion was moderated by Ricardo Henao, PhD, who kicked off the session by asking the panelists to provide an example of their experience involving AI. Suresh Balu noted that his group, the Duke Institute for Health Innovation (DIHI) typically responded to projects or problems selected by Duke Health leadership. To date, DIHI has worked with nearly 50 AI-related projects. Of these, the chronic kidney disease (CKD) model is the longest-running, but the Sepsis Watch model is the most talked-about. Balu also noted that most of the current year's (2024) projects involved large language models (LLMs) in applications including transplantation, extracting data on social determinants of health (SDOH), prior authorization for surgery and drugs, and using GPT-4 and Llama 3.1 for administrative and clinical care applications.

MODERATOR: *Given the success you've had with Sepsis Watch, do you see a shift happening? Or is there still a place for the pre-LLM models?*

Balu suggested the first response when receiving a request for an AI application should be to scrutinize whether AI is actually needed to solve a problem. If the problem can be solved without a model, it should. Otherwise, the path forward should be to pick the best possible model for the solution and put it into practice. But regardless, the choice of model, or whether to use a model at all, should be based on the merits of the case.

Dr. Jacqueline Henson, drawing on experiences working with Dr. Lisa McElroy on natural language processing (NLP) model applications for SDOH factors in transplantation, noted that despite knowing about disparities in organ transplantation for decades, clinicians and researchers have been limited in terms of being able to find causal factors. Evaluation of candidates for transplantation is done by skilled professionals who assess multiple factors, including SDOH. Dr. Henson pointed out that this data is typically found in unstructured in free-text notes in the EHR, which is where NLP comes in. Her group started with a simple algorithm and an ontology based on notes that essentially asked whether terms related to transportation were present in the chart notes, and if so, whether that signaled a barrier.

MODERATOR: *Social determinants of health have always been an issue, but natural language programs have advanced dramatically in the last few years. What do you see as driving the need for a project like this?*

Dr. Henson noted that the COVID pandemic experience, which brought to light inequities in transplantation, underscores the importance of SDOH. She further noted that transplantation is uniquely positioned to answer questions about this, partly due to regulation and because the social worker evaluation performed as part of the process typically yields rich data on SDOH.

Dr. Dan Buckland pointed out that what "AI" represents has changed over time. Currently, for many people, "AI" equals "large language model." And that for him, so far, is not really useful. Buckland went on to describe a model he *does* find useful: a neural net model that reads electrocardiograms (ECGs). Despite being a "dirty model" (i.e., not fully trustworthy) that answers a yes/no question about whether a patient is experiencing STEMI (ST-segment elevation myocardial infarction), it *is* useful at the point of care. Buckland further noted that he tends to avoid machine learning applications designed to provide data directly to doctors, suggesting that instead of making treatment decisions, machine learning might be better applied to making resource decisions, such estimating how long an operating room might be needed, or using machine learning to simulate resource shortages for planning purposes.

> *By the time a model gets to me, it works. That doesn't make it useful. I've seen a lot of models that work, but I've seen very few that are useful.*
>
> —Dan Buckland, MD, PhD

MODERATOR: *We solve AI problems with more AI. For the problem of properly evaluating a model to use in practice, do you see a way for AI to help you with that?*

Buckland pointed out that although the emergency department (ED) must make decisions on behalf of other specialties, data from the ED is not produced in ways that are useful to other specialties. Could AI help translate across these specialties, where data are structured and formatted in different ways? Large language models could be helpful in reading other clinicians' notes and translating them for use by others. Suresh Balu added that although structured data elements can be evaluated through an LLM, the final evaluation has to be done by a human. At present, LLMs seem poised to play more of a role in data processing than prediction.

Dr. Sreekanth Vemulapalli suggested that a cardiovascular algorithm that's right at least 90% of the time is at least as good as a [cardiology] fellow, adding that another potential application for AI is in diagnosis. He used the example of cardiac amyloidosis, a contributor to heart failure that is a leading cause of readmission. Cardiologists need tools that can help them recognize it, because they often miss the diagnosis when working "manually." Accurate algorithms could help cardiologists to recognize this condition and improve diagnosis and treatment of patients. A further step might be to extract data from unstructured elements and integrate it with diagnostic testing to enable a better understanding of patient risk. EHR and imaging data could be combined, and if a certain threshold is reached, that could trigger a recommendation for genetic testing.

Dr. Vemulapalli also noted that cardiovascular medicine has the advantage of significant amounts of data from large clinical trials. However, these clinical trials are very expensive: an average cardiovascular outcomes trial costs around $100M. A major component of that cost is due to the need for adjudication, in which cardiovascular experts review information to determine whether a patient did or did not have a cardiovascular event. An LLM might be able to look at that same information and render a decision.

However, to trust the output of the LLM, a range of things would be needed, including comparisons of LLM results with human evaluations and performance on historical clinical trials data, examining whether the direction or magnitude of the observed effects changed with the LLM evaluation. If those results were sufficiently convergent, parallel machine/human adjudication could be embedded in studies.

MODERATOR: *What do you see as the biggest opportunity in clinical practice? What is just hype?*

Panel participants responded to this prompt by noting:

- Risk prediction is unhelpful without an action attached to it.
- Models are great at solving the problem you *think* you have. When you talk to clinicians vs scientists vs machine learning experts, you get very different perspectives.
- In some clinical settings, risk calculation doesn't affect clinical management and may add to patient stress.
- A model by itself is just cool data science. How it fits into the clinical workflow is what's important. If the model doesn't change outcomes, what good is it?

MODERATOR: *For model development, who leads – data science or clinicians?*

Suresh Balu noted that innovation in AI is a collaborative sport. He described DIHI's approach, which includes scoping a problem in the context in of the workflow space, from patient to team to system, in order to understand the full extent of problem, then collaborating with hospitalists and nursing teams to define the solution and describe how the output will be used. Dr. Buckland added that the process of learning and working with a model goes beyond deployment and useful implementation—teams also learn how practice changes around it.

***MODERATOR**: There are numerous examples of LLMs where, if you prompt them in a certain way, you can get problematic responses. With more general use of LLMs, are they helping, or making the problem bigger?*

Dr. Henson responded to this question by noting that if clinicians don't know why an implementation is doing what it's doing, the model may just reinforce existing behavior. Suresh Balu added that one valuable feature that LLMs provide is enabling the extraction of SDOH—something that used to be difficult—and using that information to improve the model. He also noted that workflows can be changed to achieve better outcomes.

> *I'd rather have a model be consistent than "right." When variation is random, it can't be interpreted.*
>
> — Sreekanth Vemulapalli, MD

Other panelists cited other examples of flawed models, some of the pre-dating widespread use of ML/AI, such as kidney function models that excluded certain groups from being offered surgery, or pulse oximeters that did not function as well in persons with darker skin. There was general agreement that while such problems have always been present, machine learning models and LLMs could potentially amplify them because they tend to move faster.

***MODERATOR:** Is it better to have a model that is consistent but sometimes wrong, or one that has some subjectivity?*

Dr. Buckland responded by noting that "any model that gives a point estimate is wrong" and suggesting that any model should instead provide an uncertainty estimate. Dr. Vemulapalli agreed that such an approach is best practice for any model, noting that he would prefer a model be consistent rather than "right," because when variation is random, it cannot be interpreted. Suresh Balu suggested the possibility of combining multiple models, similar to the way meteorological forecasts combine multiple models to predict hurricane tracking.

***MODERATOR:** Are we holding models to different standards of evaluation than humans? If so, is it worthwhile, in terms of effort, training, and compute, to incorporate subject matter expertise?*

Balu suggested that such an approach was too expensive in terms of the training needed and that breaking problems down for agentic solutions and then recombining them was likely to prove preferable. Dr. Buckland agreed, noting that such exorbitant costs would not provide sufficient return on investment. Dr. Henson further noted that if a model incorporates current practice, it may end up reproducing existing problems. She suggested that an internal registry (Duke has one such up and running) was desirable and that information about model performance should be widely shared. A further question currently being discussed is what needs to be shared with patients: How much do patients need to know about a model being used in their care? At minimum, it would be things like model performance, any patient subgroups affected, and when the model is used in decision-making.

Experiences in Computational Medicine: The Duke Center for Computational & Digital Health Innovation

Discussants

- ***Moderator:*** *Amanda Randles, PhD; Alfred Winborne and Victoria Stover Mordecai Associate Professor of Biomedical Sciences; Director of the Duke Center for Computational and Digital Health Innovation*
- *Monica Agrawal, PhD; Assistant Professor of Computer Science and Biostatistics & Bioinformatics; Co-Founder, Layer Health*
- *David Carlson, PhD; Associate Professor of Civil and Environmental Engineering*
- *Pranam Chatterjee, PhD; Assistant Professor of Biomedical Engineering, Duke University; Co-Founder of Gameto, Inc. and UbiquiTx, Inc*
- *Jessilyn Dunn, PhD; Assistant Processor of Biomedical Engineering and Biostatistics & Bioinformatics; Director, BIG IDEAS Lab*

Dr. Amanda Randles, head of the Randles Lab and the newly formed Duke Center for Computational and Digital Health Innovation (DCCDHI) provided a big-picture overview of the evolution of work in computational medicine at Duke University and the needs that led to the creation of the DCCDHI. Dr. Randles noted that the Center had its beginnings in a series of workshops conducted at Duke that polled faculty from across multiple schools and institutes to identify grand challenges (including ones related to intramural and industry collaborations), pain points, and opportunities for innovation in computational and digital health. The goal was to identify ways that engineering expertise at Duke could be brought to bear on these issues.

The readout from those workshops identified some recurring themes:

- The existence of data and methodological silos at Duke
- Challenges in scalability of solutions
- Community-based challenges related to bridging the divides between the engineering and medical sides of campus

The Duke Center for Computational and Digital Health Innovation, which is now ramping up its programs, is dedicated to providing a multidisciplinary forum for identifying and solving these kinds of big challenges at the intersection of engineering, data science, and clinical medicine.

The Center's mission is oriented toward working with other groups at Duke, including Duke AI Health, to find disease, track disease remotely, and treat disease more effectively, focusing on enabling technologies. These include wearable devices, virtual and extended reality, and real-time computing.

Dr. Randles then introduced four faculty working with the Center, each of whom gave brief "lightning talk" presentations:

The Digital Physiome: Wearables for Disease Detection and Monitoring

Jessilyn Dunn, PhD

Dr. Jessilyn Dunn from the Pratt School's Biomedical Engineering Program introduced the work being done at the BIG IDEAS Lab (Biomedical Informatics Group – Integrate Data Engineering & Analytics). The group develops and tests tools and infrastructure using biomedical data for early detection, intervention, and prevention of disease, with much of the data coming from wearable technologies.

Dr. Dunn noted that wearables offered an appealing source for such data: not only is the data relatively inexpensive to obtain and more easily scalable, but wearables also offer seamless integration into daily life, capturing more consistent signals than measurements done in the artificial environment of the clinic. This is important because of the extent to which individual biomedical measures (e.g., heart rate) fluctuate over the course of a day. Also, because data averaged across large populations may not reflect individual variations, the granularity of wearable data enables more detailed comparisons at population levels and also better captures within-individual variation over time.

Jessilyn Dunn, PhD, Duke BIG IDEAS Lab

However, wearable technologies also present certain challenges, chief of which is the enormous volume of data that even an individual sensor can contribute over the course of day. Managing and making sense of data streams relies on the application of digital biomarkers, in which a robust relationship between digital data and a health outcome has been established.

The BIG IDEAS Lab works to develop algorithmic models that translate digital measurements to target outcomes. Similar work is underway across industry as well, but one issue that arises with marketed products is that the algorithms used to process the data may be opaque. For that reason, Dr. Dunn's group created the open-source Digital Biomarker Discovery Pipeline, which enables access to data, code, algorithms and educational resources for the entire field across academia and industry.

Dr. Dunn offered several examples of her group's work, which included the development of noninvasive proxy measures for clinically relevant measurements in diabetes. She noted that by combining the data from wearables (in this case, a smartwatch) with data from a minimally invasive device (a glucose monitor), a personalized predictive model of glucose excursions could be developed for people with diabetes, allowing highly individualized, tailored responses (e.g., diet recommendations; exercise; sleep patterns) for blood glucose control.

Another implementation example of digital biomarkers comes from the world of infectious diseases, particularly infectious respiratory diseases. During the COVID pandemic, the BIG IDEAS Lab harnessed ideas developed as part of early detection strategies in other respiratory diseases to help extend scarce testing resources during the pandemic. Using data gathered from smartwatches, the group noted characteristic changes in physical activity and heartrate that were correlated with testing positive for COVID and created an algorithm to identify people more likely to benefit from testing and waste fewer tests.

Ultimately, the goal of efforts underway through the BIG IDEAS Lab is to leverage digital biomarkers to:

- Empower patients
- Enable precision therapeutics
- Facilitate just-in-time interventions
- Improve access to care

Programmable Protein Therapeutics via Generative Language Models

Pranam Chatterjee, PhD

Dr. Pranam Chatterjee of the Pratt School of Engineering's Biomedical Engineering Program noted that his talk would be moving from the realm of disease detection to therapeutics. His work is centered on the development of programmable proteins to treat disease, integrating algorithms to create proteins that can be used to engineer other proteins and even cells and tissues.

Pranam Chatterjee, PhD, Department of Biomedical Engineering

Dr. Chatterjee pointed out that certain kinds of proteins known as transcription factors (TFs) are already being used to engineer different types of cells: when introduced to stem cells, transcription factors can transform the stem cells into other desired types of cells, such cardiomyocytes or adipocytes. However, some types of cells, such as oocytes and other germ cells, have proven difficult to engineer with the combination of stem cells and TFs.

Working with graph theory algorithms, Dr. Chatterjee's group has developed methods for identifying candidate programs that can then be screened in the lab and sequenced to identify three TF proteins that produced oocyte-like cells. One application of this technology is being used by a company called Gameto that was founded by Dr. Chatterjee, in which immature oocytes are successfully cultured.

In addition, proteins can be used to edit DNA. CRISPR Cas-9 is one well-known example, but CRISPR technology is only able to edit a fraction of all DNA sequences. Dr. Chatterjee's

group has created a series of algorithms to discover and tailor new enzymes that expand the reach of CRISPR-like technologies that can be applied to a host of genetic diseases.

Another application relies on designed peptides, called ubiquibodies, that allow for selective degradation of proteins, including viral proteins. Most disease-causing proteins are "undruggable" with small-molecule drugs. However, Dr. Chatterjee's lab uses protein language models, a variation on natural language models, to design peptides in combination with image generators that illuminate the structure and function of proteins. A further model that uses language models that can be given a protein target for a peptide will return a candidate molecule with 38% accuracy. The system can also be used to reverse the degradation of and stabilize desirable proteins.

Scalable Natural Language Processing to Transform Healthcare

Monica Agrawal, PhD

Dr. Agrawal described her work, which bridges both computer science and medicine and is focused on realizing the potential for data contained in electronic health records (EHRs) to enable personalized approaches to medicine. Dr. Agrawal pointed out that conventional clinical trials often cannot answer questions that would be useful for guiding medical practice and clinical decision-making, such as:

- What drug would lead to the best outcome for *this* patient?
- What is the patient's expected disease trajectory?
- What adverse events might emerge from drug combinations?

Monica Agrawal, PhD, Department of Biostatistics & Bioinformatics

However, answering these questions requires access to granular data. For the example of predicting what adverse events might emerge from a given set of drug combinations, the variables needed for outcomes research would include items such as:

- The patient's disease
- Disease status or progression
- Interventions or therapies attempted
- Symptoms or side effects experienced
- Confounders, such as comorbid conditions

Although these data can be found in the EHR, they are inherently "messy" – there is no single table that houses these variables. Due to splintered care and deviations from original care plans, existing formal documentation cannot be assumed to be correct, and changes or updates are often found in free-text clinical notes written or dictated by the clinician.

Despite the potentially rich information contained in chart notes, accurately parsing what they mean requires both general familiarity with medicine and a detailed grasp of the

context, which may be critical for understanding abbreviations and acronyms, many of which may have multiple interpretations.

Until now, the status quo for extracting useful information from unstructured data and transforming it into structured data that is useful for research required time- and labor-intensive chart review by human experts, or a combination of chart review and the application of a machine learning/NLP model. However, even with the use of partial chart review plus machine learning, the process still represents a major bottleneck: each variable of interest requires the creation of a new model, and chart review and annotation still take substantial amounts of time. Further, sharing clinical notes can be difficult to share across institutions due to the amount of sensitive and potentially identifiable information they contain.

With the arrival of LLMs such as the various GPTs, this process can be made more efficient and scalable. They can be used for tasks, such as parsing clinical notes, extracting acronyms and medical information that previously required separate models and training data. However, naïve use of LLMs for this purpose can create problems. Techniques such as weak supervision can help overcome this, boosting LLM performance and improving deployability.

LLMs also allow the process of clinical documentation to be reimagined. In one pilot study, a contextualized autocomplete program, personalized for individual patients, automatically normalizes concepts to ontologies while the chart note is being written. The tool decreases clinicians' documentation burden and reduces the number of keystroke errors. Information retrieval was also streamlined, so that typing a particular condition (for example, "afib" normalized to "atrial fibrillation") opens a window that contains a list of medications for that condition, saving time and reducing number of clicks necessary.

Looking to the future, new workflow applications for LLMs include:

- Answering questions for patients
- Ambient clinical documentation
- Record and note summarization
- Translation of clinical jargon to patient-friendly language

However, there are also attendant challenges arising from the use of LLMs, such as:

- Hallucinations and omissions
- Agreement on a "gold standard"
- Evaluating free text
- The contemporaneous evolution of clinical knowledge
- Inadequate or flawed data used to train LLMs
- Excessive trust or over-reliance on LLM output

Language is embedded across medicine and affects stakeholders across medicine. The goal is to make medical language work for everyone.

Machine Learning to Infer & Modify Brain State

David Carlson, PhD

Dr. Carlson, an associate professor in the Departments of Civil and Environmental Engineering and Biostatistics and Bioinformatics at Duke, focused his presentation on the idea of building new abilities in applying technology, specifically using machine learning to infer brain states and even to control them.

What is a brain state? Using the example of a mouse, Carlson explained that the basic idea is to collect data on the functioning of an individual brain and to transform that data into knowledge about how the animal is feeling (e.g., angry, anxious, aggressive, fearful, happy, contented). If it is possible to understand whether a mouse is feeling threatened or aggressive—and therefore likely to attack other mice or behave aggressively toward them—it might also be possible to intervene and prevent that behavior by changing the brain state.

While acknowledging that the premise might sound somewhat science-fictional, Carlson described the use of machine learning to map brain network models of emotions and behaviors, noting that such models must also be able to explain how it makes such predictions to the neuroscientists using it. Ultimately, the goal is for this *explainable AI* to generate testable hypotheses and design confirmatory experiments and interventions.

David Carlson, PhD, Department of Civil and Environmental Engineering

Brain data from test subject animals are used to create a map, called an electrical functional connectome, that identifies the different regions of the brain (and the connections between them) that are engaged when an animal is experiencing an emotion such as stress. If such patterns can be interpreted accurately, the behavior can be modified through neurostimulation or pharmaceutical interventions.

Dr. Carson introduced an experimental assay comprising test cases in which a male mouse is placed in proximity with other mice that could trigger either aggressive or nonaggressive behaviors depending on circumstances. These experiments revealed brain networks that, when expressed, were associated with either aggressive or peaceable behavior.

The next stage of the investigation focused on whether these machine-learning-derived brain network states could be used to develop ways to affect behaviors such as aggression. One way to accomplish is through a technique called optogenetic stimulation, in which proteins that are sensitive to light can be triggered to cause particular neurons in a particular region of the brain to activate or switch off. One interesting outcome of the study, which was adapted from a prior investigation of optogenetic switching in humans, was the finding that continuous stimulation of parts of the brain known to reduce aggression not only yielded that (expected) result but could also be done without affecting prosocial behaviors. What's more, the machine learning model revealed a very specific pathway

implicated in aggression that could permit intervention on aggression while also enhancing prosocial behaviors.

In his conclusion, Dr. Carlson noted ongoing collaborative investigations in contexts such as anxiety and depression.

Computational Medicine Interactive Panel

Hosted by the Duke Center for Computational and Digital Health Innovation

- ***Moderator:*** *Amanda Randles, PhD; Alfred Winborne and Victoria Stover Mordecai Associate Professor of Biomedical Sciences, Director of the Duke Center for Computational and Digital Health Innovation David Carlson, PhD; Associate Professor of Civil and Environmental Engineering*
- *Karnika Singh; PhD Candidate at Duke University; Member of the BIG IDEAS Lab*
- *Cyrus Tanade, MEng; PhD Candidate at Duke University and NSF GRFP Fellow; Member of the Randles Lab*
- *Monica Agrawal, PhD; Assistant Professor of Computer Science and Biostatistics & Bioinformatics; Co-Founder, Layer Health*

Panel Chair Amanda Randles, PhD, Director of the Duke Center for Computational and Digital Health Innovation, moderated the discussion. Following a brief introduction of the participants that also touched on the recent awarding of Nobel Prizes related to AI, she posed a series of questions to panel participants related to the promise and challenges of computational science in biomedical research and clinical practice from the data science perspective.

MODERATOR: *What do you see as being the barriers to innovation – is it methods, data, regulations?*

Dr. Agrawal noted that while innovation is a relatively small part of the translation equation, it is nonetheless what academia tends to focus on. Meanwhile, data distribution shifts and data silos affect how well researchers can generalize, whether between hospitals or across the same hospital over time. She also observed that the COVID pandemic broke many models because of changes to this landscape. Regulation also presents challenges, with the FDA's regulatory model being based on AI considered as a medical device. But classical medical devices tend to be more static in nature, whereas models are constantly trained on new data. This means that the current regulatory model is not based on a suitable paradigm, even in the case of a predictive model with a simple yes-or-no output. Finally, Dr. Agrawal pointed out that the output from an LLM currently cannot be compared with the reasoning and judgment of a real doctor, and that the overarching challenge is to learn how to build models that will generalize and find ways to evaluate them effectively.

***MODERATOR**: Graduate students are the "boots on the ground." What challenges are you encountering in getting training and compute resources needed to complete a PhD and still build something that will make a difference?*

Cyrus Tanade discussed data challenges related to the use of physics-based models, in particular the assumption is that supplying the model with increasing amounts of patient-specific data at it will inevitably yield a better model, which is not always the case. Instead, Tanade suggested that this assumption needs to be flipped on its head: what if all these parameters are not the most important for the prediction of interest – in his particular case, modeling blood flow? From 14 measurements, physics-based modeling needed just two. Rather than incorporating enormous amounts of data that would involve impracticable amounts of compute time, innovative solutions, such as breaking down time series into puzzles, can offer better solutions.

Karnika Singh suggested that generalizability poses the biggest challenge, and that models are generally as good as data that's fed into them. The challenge lies in ensuring that training data comes from diverse groups that represent the population(s) in which the model will be deployed. She noted that studies conducted during the COVID pandemic that relied on a bring-your-own-device data plan revealed large demographic imbalances. However, in attempting to recruit more diverse and representative participants, the amount of time needed to participate in the study was revealed as a potential problem, one that was helped by incorporating a completely remote study design.

***MODERATOR**: Can you talk a little about the entrepreneurial journey?*

Dr. Agrawal described her positive experience with the health tech startup world, noting that although the world of academic computer science was highly productive in terms of publications and conference presentations, there remained a need for translation to the real world. Examples of this include the lack of capability for EHRs to incorporate longitudinal data from wearable devices, challenges related to privacy and interoperability, and dealing with messy and incomplete data.

***MODERATOR:** Patients are increasingly empowered, but that's a shift from when everything was in the EHR and was coming from hospital. How we evaluate and incorporate these new data sources is a challenge but also a good opportunity. What areas are ripe for innovation over the next 5-10 years?*

Dr. Agrawal suggested that one area ripe for innovation lies in the integration of different modalities and building models. Dr. Carlson agreed, noting that the challenge of integrating multiple data streams and improving analysis by using existing data permeates every scientific field. Carlson pointed to toxicology as an example: there is a lack of definitive information about how safe many substances are, but integrating consumer data and feedback would yield a better understanding of the toxicity of consumer products.

***MODERATOR:** What are the biggest misconceptions about what AI can do in health?*

Participants offered up a range of misconceptions, including:

- Enabling self-diagnosis: most AIs are not trained on medical data.

- That AIs can outperform doctors: most models' performance in this regard is based on a very narrow set of questions (such as those in USMLE) that don't capture the fullness of medical practice.

MODERATOR: *AI/ML instruction is very popular on Coursera. What is the role of the university here? What gaps can it fill?*

Cyrus Tanade suggested that one key benefit that universities offer is a collegial atmosphere that's helpful in getting up to speed and keeping up. Karnika Singh agreed, citing the "ecosystem" at Duke, which includes access to people performing research, journal clubs, and the "real-world" environment of Duke Health immediately at hand. However, Dr. Carlson noted that one thing that could be improved upon is the depth and breadth of curriculum, suggesting that all university graduates, and perhaps all modern citizens, need a degree of basic computer and data literacy.

MODERATOR: *What advice do you have for balancing needs for multiple skillsets? How do you bridge interdisciplinary teams?*

Dr. Agrawal cited clinical shadowing as useful in this regard, as it allows one to see how software is actually used, and how data are generated and retrieved. She observed that it's helpful to understand that ICD codes are based on billing rather than treatment—something that lets you ask yourself whether you're addressing the correct problem. Acquiring a basic medical-clinical vocabulary to communicate with clinicians was also helpful. Dr. Carlson agreed, adding that not being afraid to ask "stupid" questions was likewise helpful.

Thematic Wrap-up

On the final morning of the Duke Summit on AI for Health Innovation, summit participants convened for a summary and read-out of key themes that had emerged over the course of multiple presentations and discussion panels, summarized and grouped here according to major themes:

Adoption & Implementation

- Understanding workflow and personnel is critical for successful adoption.
- Layering technology solutions on top of flawed/chaotic systems will not help and will likely hurt.
- Continuous evaluation and transparency are important components for building trust in AI/ML technologies.

Cultural Considerations

- AI can achieve savings not by replacing people, but by eliminating waste and defects.
- Convening cross-disciplinary expertise is often vital for success, but everyone has to learn a common vernacular to communicate effectively.
- Collegial communication holds disparate teams together and keeps them focused on tasks.
- Individuals and communities affected by AI have to be part of design and implementation decisions.

Major Caveats

- Just because it works doesn't mean it's useful.
- Models are not useful unless their output informs or enables action.
- Matching the application to the workflow is critical for success.
- Model consistency may trump model "correctness."
- The persistence of silos means that data collected in one location, treatment context, or population may not be consumable in another.
- AI may have potential for translating data gathered in one setting for application in a different one.

Current & Emerging Challenges

- Acting on opportunities for progress (beyond "low-hanging fruit" of data extraction).
- Bridging academia and industry; creating pathways for successful translation and product development.
- Coping with silos in and beyond healthcare.
- Integrating data on social determinants of health and continuously monitoring model performance.
- Incorporating measures of confidence/uncertainty into model output.

Supplemental Materials

Industry Panel

The programming for the Health AI Summit included a panel discussion focused on the perspectives of expert representatives from the healthcare and health research industries, who were invited to share their thoughts and experiences related to adapting AI tools for healthcare.

The panel, moderated by Jonathan Owens, director of industry engagement for Duke Engineering, included:

- Basia Coulter (Globant Healthcare & Life Sciences)
- Brendan Fowkes (IBM)
- Marissa Halfin (LabCorp)
- Kirsten Lum (Johnson & Johnson)
- Terry Myerson (Truveta)

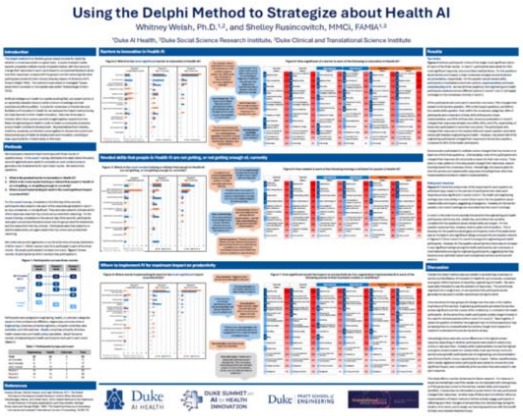
The Pratt School of Engineering's Ken Kingerly covered this wide-ranging discussion in an article available at the Pratt website:

- **Industry Insights into Developing AI-Powered Healthcare Solutions**

Summit Wrap-Up: Delphi Survey Results

To probe for consensus on the barriers and facilitators of innovation in Health AI, we invited Summit attendees to participate in a classical Delphi study with three rounds of questionnaires. We asked three questions:

1. What is the greatest barrier to innovation in Health AI?
2. Which is the most needed training or skillset that people in Health AI are not getting, or not getting enough of, currently?
3. Where would implementing AI result in the most significant impact on productivity?



For a summary of survey results, insights, and feedback prepared by Duke AI Health Research Scientist Whitney Welsh, PhD, please visit: https://duke.box.com/s/h3cbyadtyltpqdysiadk9g3fdrdjellw

In addition to the survey results summary, a poster, authored by Dr. Welsh and by Duke AI Health Managing Director Shelley Rusincovitch, MMCi, describing the survey methods and results, was presented at the 2025 Association for Clinical and Translational Science in

Washington, DC. An accompanying abstract was published in a supplementary issue of the *Journal of Clinical and Translational Science*:

> *Welsh W, Rusincovitch S. 379 Using the Delphi method to strategize about health AI. J Clin Transl Sci. 2025 Apr 11;9(Suppl 1):117. doi: 10.1017/cts.2024.1002. PMCID: PMC12050691.*