



Case Study

Array Insights

Written by Audrey Woods

There's a lot of enthusiasm about the transformative potential of Artificial Intelligence (AI) in healthcare. Already AI programs are being implemented to help diagnostics, treatment decisions, drug design, genetic studies, and wide swaths of medical research. However, for these AI services to be robust and trustworthy, they must be trained on huge volumes of medical data, data that can be difficult to access. Hospitals are hesitant to share their data and when they do, historical methods have left patients open to the risk of a privacy breach.

MIT spinoff and CSAIL Alliances Startup Connect member [Array Insights](#) is tackling this problem with a combination of privacy-preserving technologies, offering researchers a unique platform that allows them to access diverse medical data in a way that keeps patient information secure.

GETTING THEIR START

CEO and co-founder Anne Kim explains how the idea for Array Insights started when she and other postdocs at CSAIL were conducting research that required patient data from clinical trials, gene association studies, and hospital intensive care units. She and her colleagues were shocked at how this sensitive and private information was handled, with outdated file transfer protocols and even physical hard drives shipped in the mail. She says they were getting copies of data in a way that gave the providers “no ability to track and trace the usage of it.” So in 2017, Kim and fellow researchers, along with [CSAIL Professor Manolis Kellis](#), set off to tackle the question: “how do we give patients the ability to control the usage of their data?”

The fundamental premise, Kim says, could be boiled down to “23andMe plus blockchain.” The Array Insights founders believed that if patients had essentially an NFT of their genetic data, it would give them the ability to manage their information while simultaneously offering researchers better access to diverse medical data. As Professor Kellis explained in a 2021 [profile](#) of the company, “you shouldn't have to schmooze with hospital executives for five years before you can run your machine learning algorithm... **[Array Insights'] goal is to help patients, to help machine learning scientists, and to create new therapeutics. We want new algorithms—the best algorithms—to be applied to the biggest possible dataset.**”

However, Kim describes how the company faced early headwinds when marketing directly to patients. She says Array Insights bumped against an activation energy required to get patients engaged with controlling their own data, one which proved to be “a really difficult challenge.” This inspired the company to pivot, pitching instead to pharmaceutical companies and hospitals. This was better, but “really slow,” Kim says, as they were spending a lot of time matchmaking between pharmaceutical companies who wanted specific data and hospitals who had that data, and patients were still left out of the loop. At that point Kim and her colleagues began to investigate how to get back to Array Insights' original concept in a way that kept alive the beating heart of the company: patient empowerment.

This led them to begin working with patient advocacy groups. Kim explains how there are 10,000 patient advocacy groups in the US alone and their fundamental mission aligns with that of Array Insights: to advocate for patient rights. These advocacy groups also manage patient data in enormous and disease-specific registries, which was a perfect fit for their needs. Associating with patient advocacy groups offered benefit all around, as Array Insights could gather diverse research data, the patient advocacy groups could participate in groundbreaking research without risking confidentiality, and patients remained protected.

A NEW WAY TO TRAIN MEDICAL AI MODELS

The main product of Array Insights is their Unified Patient Registry, which gives researchers access to clinical datasets from hundreds of disparate hospitals. The technology underlying this interface is a combination of two security concepts in computer science: federated learning and secure enclaves.

With federated learning, the idea is that instead of sharing the data with a researcher the old-fashioned way—risking privacy and the potential for malicious activity—the program sends the machine learning model to the source of the data where it is trained locally, updated, and then sent back to the researcher. This keeps all sensitive data on, for example, a hospital’s server, which gives the data provider confidence while offering researchers the opportunity to train their models.

Still, Array Insights needed to solve the other half of the problem, which was giving researchers the confidence to share proprietary models with hospital systems without risking that such models might be copied or stolen. For this, Array Insights applied secure enclaves, also known as Trusted Execution Environments (TEEs). TEEs—the same technology that makes entertainment companies like Netflix and Spotify possible—are a hardware-backed way to trust that the computation at the end user level is doing what the programmer intended it to do. For example, TEEs allow Netflix to know that a user is watching a movie but not able to copy it. With Array Insights, TEEs “provide access to a library of patient records for the sake of utilizing them but also keeping them private.” Kim says applying these two security measures gives users “trust in the code, trust in the hardware, and trust in the environment.”

CONNECTING THROUGH CSAIL: FUTURE LABS CAPITAL & MORE

Kim describes the experience of creating a startup as “a marathon and also a sprint.” But she says that this roller-coaster process has been helped tremendously by MIT resources such as CSAIL Alliances. She credits Alliances with “being able to connect us with a bunch of different resources,” saying **“the fact that we were able to get in front of Novo Nordisk and other pharmaceutical companies [was] very important for product market fit and discovery.”**

Of course, fundraising is one of the biggest priorities of new startups, which is why Kim is grateful CSAIL Alliances also connected Array Insights to Future Labs Capital who she says have been “really great investors.” She describes how beyond funding them, Future Labs Capital “has been helpful and active in being able to say: how can we help you next?” which Kim says has contributed to Array Insights’ success.

When speaking about her relationship with CSAIL Alliances, Kim goes out of her way to highlight specific members of the Alliances team who have helped both her personally and Array Insights as a whole. Of Client Relations Manager Philip Arsenault, Kim says, “he just cares about people and knows how to get things done.” Generally, she credits CSAIL and CSAIL Alliances for creating a community of entrepreneurship and positive encouragement that has made her motivated to support other students and give back in any way she can.

For more information about CSAIL Alliances industry engagements, please visit:

cap.csail.mit.edu

LOOKING FORWARD

As important as patient advocacy groups have been for Array Insights, Kim says that she and her colleagues are excited to get back to their original idea of connecting directly with patients. They've recently started meeting and engaging with individual patients in a process that Kim describes as "really energizing."

When it comes to CSAIL Alliances, Kim is excited for future events and opportunities to get their product in front of companies, which will allow the startup to continue growing and expanding. "It's always exciting," she says of their current plans, expressing how she hopes the company's "learning growth mindset" will continue the momentum of their success.

Overall, Kim says she's grateful to CSAIL and to the support she's received from programs such as CSAIL Alliances, concluding with the sentiment: **"I love MIT."**