White Paper

Health and Life Sciences
Artificial Intelligence

# Privacy-Preserving Data-Collaboration Methods that Accelerate Healthcare Innovation

## Confidential computing platforms and privacy-preserving analytics will help healthcare organizations accelerate the development and use of clinical algorithms

**Authors**

**Chris Gough**
Worldwide General Manager, Health & Life Sciences
Intel Corporation

**MaryBeth Chalk**
Co-founder and Chief Commercial Officer
BeeKeeperAI, Inc.

In the continued quest to improve patient outcomes and lower costs, healthcare organizations (HCOs) are looking to technology, particularly advances in the field of artificial intelligence (AI) to spur exciting innovations. Such innovations have the potential to help with disease prediction and diagnosis, effective treatment selection and prognosis, life sciences and pharmaceutical research, epidemiology, public health, and precision health initiatives.

While these approaches hold great promise to fuel future breakthroughs in healthcare and care delivery, they require access to sufficient quantities of diverse data for the development and validation of models capable of consistent performance. Thanks to electronic health records (EHRs), medical devices and personal smart devices, as well as the data collected in groundbreaking research studies at different academic medical centers around the globe, more and more data is available than ever before. The problem, however, lies in how to safely and ethically access, integrate, and then analyze the information while preserving individual privacy.

“The healthcare AI market is projected to be a $46 billion market within the next five years,” said MaryBeth Chalk, Co-founder and Chief Commercial Officer at BeeKeeperAI, Inc. “But today, there are approximately 30 algorithms that have been approved by the FDA for clinical use – and only two of those are considered *de novo*, or truly unique. To get to truly novel clinical breakthroughs, we need new solutions to overcome the traditional barriers to accessing the real-world data required for generalizable algorithms.”

As noted in *The American Journal of Medicine*, HCOs require a privacy-preserving framework to support the kind of data collaboration that will make tomorrow’s scientific and clinical advances possible – while supporting improved patient outcomes and experiences.[1] Unfortunately, establishing this sort of framework has historically been a challenge.

“Within the current environment, the data needs to be protected in transit, at rest, and in use – and it’s very complex, if not impossible, to provide that level of security,” said Chris Gough, Worldwide General Manager, Health & Life Sciences, Intel. “But even when it’s done, a major gap that still exists is the intellectual property [IP] protection of any AI model accessing the data. As more companies and start-ups work in this area, whose value will lie on the IP of their particular models, this is a huge concern. There needs to be a way to enable organizations to collaborate while maintaining confidentiality and security.”

## Creating more generalizable AI models

To create successful algorithms for use at the bedside, developers need to find ways to make their models generalizable to diverse populations.

"As we go through regulatory approval for the algorithms, we need to meet this standard of generalizability," said Chalk. "Simply put, it means that your algorithm works no matter what equipment you use, no matter what ethnicities or races you serve, no matter what clinical workflows you employ, and no matter where in the world your algorithm is used. No matter the circumstances, your algorithm is only generalizable if it performs consistently, with the same level of accuracy, regardless of the setting. The only way you can achieve this this level of consistent performance is if you have trained and validated your algorithm on all the different variables – and that requires access to highly diverse, real-world clinical data."

Traditionally, to train and validate AI models, collaborating HCOs would have to go through a complex and time-consuming process. Gough said two hospitals, for example, would have to first agree on standards for data – including de-identification, semantics and nomenclature, just to start – and then decide where to move the data so it could be stored and used.

"There are a lot of obstacles with this approach," he explained. "First and foremost, standards may differ depending on what country your organization is in. In the United States, we have the Health Insurance Portability and Accountability Act [HIPAA]. The General Data Protection Regulation [GDPR] is more of a global standard. Either way, patient privacy needs to be protected – and in some countries, privacy laws may forbid any patient data from ever leaving the country. That means organizations outside of that country won't have access to it."

But such agreements are only the beginning of the data-sharing process. HCOs also need to consider how to move, replicate, and store datasets – which can be quite costly. They also need to find ways, even as cyberattacks are increasing across the healthcare ecosystem, to keep that data, as well as the proprietary algorithms using it, secure.

"You are increasing the attack surface for potential cyberattacks when you've got multiple copies of data in multiple places, to start," he said. "But then you have a lot of other complexity to manage. How do you keep the different datasets in sync? Do you need to de-identify the data, which can be quite difficult to do with unstructured data? And that's all before you consider the IP concerns surrounding new algorithms."

Addressing these different concerns requires significant time and resources. Chalk said today's data-sharing processes can take anywhere from 36 months to five years at a cost of $3-5 million to validate just a single AI model.

"You can see why getting access to sufficient amounts of real-world data, especially rare or low prevalence diseases, has been virtually impossible," she said. "We need a way to cut down on the cycle time so we can generate more novel solutions for use in healthcare delivery – and get them out to market faster."

## A secure way to support data collaboration

Confidential computing platforms (CCPs), with memory encryption and privacy-preserving analytics, however, support HCOs in overcoming many of those traditional hurdles by helping protect data at rest and data in use. In fact, these platforms, Gough said, don't move or share data, per se – rather, they enable data collaboration through trusted and secure execution environments. And while such platforms came of age in the financial services and governmental security industries, they offer many benefits to HCOs – enabling industry leaders to rethink the ways they can leverage data for advancing healthcare outcomes and quality.

"These platforms enable more secure enclaves where only approved software applications and approved data can be operated on – neither the provider of the data nor the provider of the application or algorithm has visibility into what's happening inside the environment," he said. "They can only receive the derived data that comes from that processing. In a sense, this isn't really data sharing because the data doesn't move and it isn't visible. It is data collaboration that allows the encrypted data to be operated on in the HCO's secure environment.

This high-tech collaboration is made possible by **Intel® Software Guard Extensions (Intel® SGX)**. While Intel SGX has been in production for a number of years, the much larger enclave sizes supported by **3rd Gen Intel® Xeon® Scalable Processors** make this compute solution a great match for large data AI/analytic workloads.

Chalk added that the use of CCP transforms the traditional model of accessing and aggregating data, leading to tangible benefits for everyone. "[Confidential computing platforms] allow us to reduce the cycle time to validate an algorithm in half," she said. "It also cuts the costs almost in half. Those kinds of savings allow us train, validate, and bring to market generalizable algorithms much faster. And, it will only get faster and less costly as the technology and processes underlying CCP mature."

BeeKeeperAI has worked to validate three different clinical models using an Intel-based CCP, including a hemodynamic stability index, a COVID-19 detection tool, and a treatment stratification tool for diabetic retinopathy. But, she maintained, the possibilities for different clinical algorithms are endless.

"These models use all sorts of different data – including EHR, lung X-rays, and fundus images of the eye – without difficulty," she said. "The data doesn't move anywhere. The data stays in the secure environments where it already exists. BeeKeeperAI provides a secure way for algorithm owners to compute on the real-world data they need to achieve generalizability while the data remains in control of the data steward at the originating institution."

Both Gough and Chalk hope that technology stakeholders understand that the use of such CCPs allow a variety of different data collaboration use cases that simply weren't feasible before. And, in doing so, they can help propel the sorts of AI-driven advances that can make significant improvements to care delivery.

"When you don't have to move data, you don't need to take the time to de-identify it," said Gough. "Any data sovereignty issues can be addressed because the data never needs to leave the geographic boundary of that country, state or province, institution, or even department. It permits a new class of multi-party analytic workloads that will fuel tomorrow's clinical innovations."

Privacy-preserving, multi-party analytics will provide a strong foundation for advances in diagnostics, prognostics, and population health initiatives in the future. Powerful distributed computing platforms that can support new AI algorithms have the power to lower healthcare costs, while empowering clinicians with the information they need to improve both patient outcomes and patient experiences. And it is clear they have the power to help guide tomorrow's most groundbreaking clinical applications.

"In order to fulfill the mission of healthcare organizations – particularly the academic medical centers – you need to be able to leverage your data for scientific discovery," said Chalk. "The traditional means of accessing, sharing, and using data has been a major barrier to scientific discovery. The opportunity of the future is to embrace technologies that allow us to do arms-length computing in a more secure way, where the technology itself ensures trust among parties in which there is no exposure of private or proprietary information. This is a fundamental disruption of scientific and clinical collaboration."

To learn more about Intel's vision for the future of healthcare, visit:
**intel.com/health**

[1] Noorbakhsh-Sabet, N., Zand, R., Zhang, Y., and Abedi, V. 2019. Artificial intelligence transforms the future of healthcare. The American Journal of Medicine 132(7): 795-801.

Intel technologies may require enabled hardware, software or service activation.
No product or component can be absolutely secure.
Your costs and results may vary.
Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.

0721/EOH/MESH/PDF 347708-001US