Accelerating healthcare AI innovation with Zero Trust technology

Posted on October 26, 2021

John Doyle Chief Technology Officer, Microsoft Health & Life Sciences

From research to diagnosis to treatment, AI has the potential to improve outcomes for some treatments by 30 to 40 percent and reduce costs by up to 50 percent. Although healthcare algorithms are predicted to represent a $42.5B market by 2026, less than 35 algorithms have been approved by the FDA, and only two of those are classified as truly novel.[1] Obtaining the large data sets necessary for generalizability, transparency, and reducing bias has historically been difficult and time-consuming, due in large part to regulatory restrictions enacted to protect patient data privacy. That's why the University of California, San Francisco (UCSF) collaborated with Microsoft, Fortanix, and Intel to create BeeKeeperAI. It enables secure collaboration between algorithm owners and data stewards (for example, healthy systems, etc.) in a Zero Trust environment (enabled by Azure Confidential Computing), protecting the algorithm intellectual property (IP) and the data in ways that eliminate the need to de-identify or anonymize Protected Health Information (PHI)—because the data is never visible or exposed.

Enabling better healthcare with AI

By uncovering powerful insights in vast amounts of information, AI and machine learning can help healthcare providers to improve care, increase efficiency, and reduce costs. For example:

- AI analysis of chest x-rays predicted the progression of critical illness in COVID-19 patients with a high degree of accuracy.[2]
- An image-based deep learning model developed at MIT can predict breast cancer up to five years in advance.[3]
- An algorithm developed at the University of California, San Francisco can detect pneumothorax (collapsed lung) from CT scans, helping prioritize and treat patients with this life-threatening condition—the first algorithm embedded in a medical device to achieve FDA approval.[4]

At the same time, the adoption of clinical AI has been slow. More than 12,000 life-science papers described AI and machine learning in 2019 alone.[5] Yet the U.S. Food and Drug Administration (FDA) has only approved a little over 30 AI- and machine learning-based medical technologies to date.[6] Data access is a major barrier to clinical approval. The FDA requires proof that a model is generalizable, which means that it will perform consistently regardless of patients, environments, or equipment. This standard requires access to highly diverse, real-world data so that the algorithm can train against all the variables it will face in the real world. However, privacy protections and security concerns make such data difficult to access.

Breaking through barriers to model approval

As both an AI innovator and a healthcare data steward, UCSF wanted to break through these challenges. "We needed to find a way that allowed data owners and algorithm developers to share so we could develop bigger data sets, more representative data sets, as well as allowing [data owners] to get exposed to algorithm developers without risking the privacy of the data," says Dr. Michael Blum, Executive Director of the Center for Digital Health Innovation (CDHI) at UCSF.[7]

With support from Microsoft, Intel, and Fortanix, UCSF created a platform called BeeKeeperAI. It allows data stewards and algorithm developers to securely collaborate in ways that provide access to real-world, highly diverse data sets from multiple institutions, where AI models are validated and tested without moving or sharing the data or revealing the algorithm. The result is a Zero Trust environment that can dramatically accelerate the development and approval of clinical AI.

BeeKeeperAI relies on a unique combination of software and hardware available through Azure Confidential Computing. The solution uses virtual machines (VMs) running on specialized Intel processors with Intel Software Guard Extensions (SGX). Intel SGX creates secured portions of the hardware's processor and memory known as "enclaves," encrypting and isolating the code and data inside. Software from Fortanix handles encryption, key management, and workflows.

Proving the Zero Trust model

In June of 2021, the BeeKeeperAI platform demonstrated the ability to send algorithm models via the Azure Confidential Computing environment to two data steward environments. Upon verification, the model and the data entered the Intel SGX secure enclave, where the model was able to validate against the PHI data sets. Throughout the process, the algorithm owner could not see the data, the data steward could not see the algorithm model, and BeeKeeperAI could see neither the data nor the model. The platform completed and passed a third-party HIPAA security audit and the first product release, EscrowAI, will be commercially available on Azure Marketplace in March of 2022.

BeeKeeperAI is currently working with aiScreenings, a Microsoft partner headquartered in France, to demonstrate the platform's global applicability as it facilitates the validation of aiScreenings algorithm for identifying retinopathy. "A critical advantage of BeeKeeperAI's Zero Trust environment is its compliance with EU General Data Protection Regulation data security standards," says Arnaud Lambert, CEO of aiScreenings. "BeeKeeperAI accelerates our time to market by reducing the effort we have historically spent to verify the performance of our algorithms against U.S. patient data." aiScreenings plans to use BeeKeeperAI to evaluate algorithms for critical cancer-based pathologies.

Collaborating for better care

This is only one example of how improved access to multi-site and rare data sets will open opportunities to develop novel algorithms that can improve care, reduce costs, and save lives. Additionally, the BeeKeeperAI team estimates that its technology may be able to reduce time to market by as much as 12 months and save $1M to $3M in development costs for a typical project.

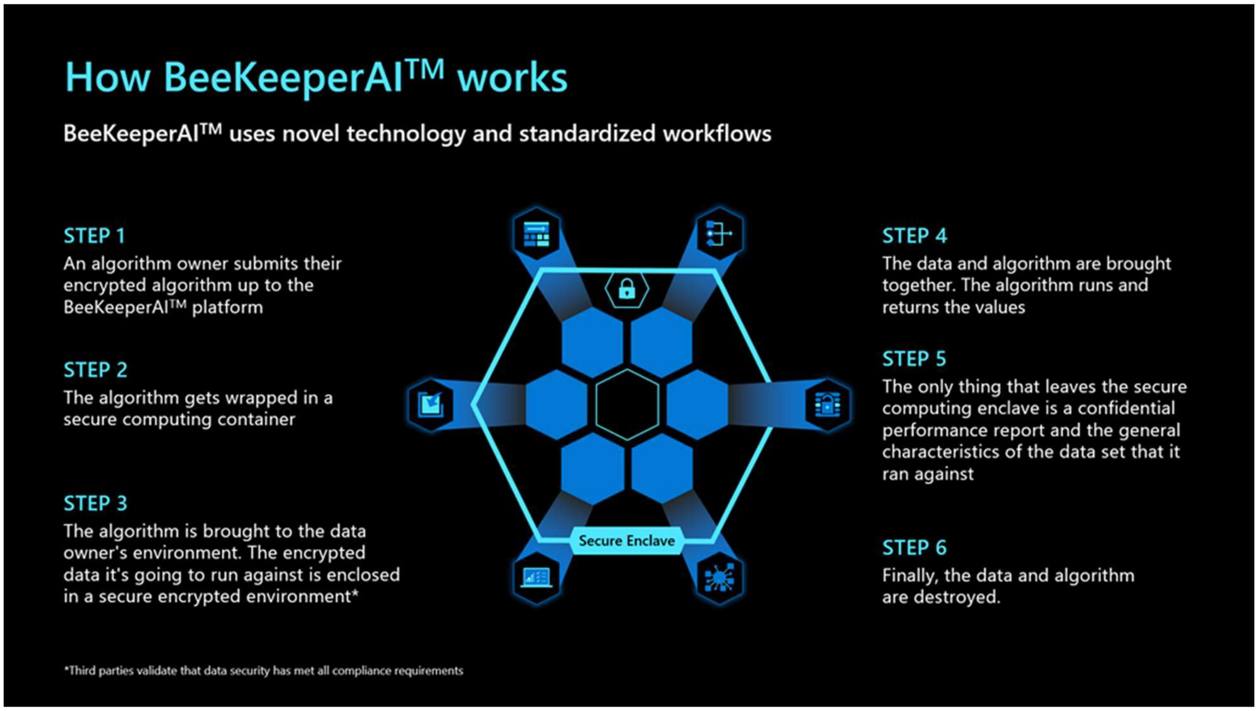
"Microsoft has invested heavily in creating tools for healthcare and enabled BeeKeeperAI to assemble the capabilities required for a Zero Trust platform that will be deployed directly from the Azure marketplace," says Bob Rogers, Ph.D., co-inventor and co-founder of BeeKeeperAI and Expert in Residence for Artificial Intelligence (AI) at UCSF's Center for Digital Health Innovation. "Additionally, Azure is the only cloud we could use to access Intel SGX technology, which is a critical component of our Zero Trust platform."

"When researchers create innovative algorithms that can improve patient outcomes, we want them to be able to have cloud infrastructure they can count on to achieve this goal and protect the privacy of personal data," says Scott Woodgate, Senior Director, Azure Security and Management at Microsoft. "Microsoft is proud to be associated with such an important project and provide the Azure confidential computing infrastructure to healthcare organizations globally."[8]

Bringing together hardware and software security

The data steward uploads encrypted data to their cloud environment using an encrypted connection that terminates inside an Intel SGX-secured enclave. Then, the algorithm developer submits an encrypted, containerized AI model which also terminates into an Intel SGX-secured enclave. The Key Management System enables the containers to authenticate and then run the model on the data within the enclave. The data steward never sees the algorithm inside the container and the data is never visible to the algorithm developer. Neither component leaves the enclave. Even a malicious admin or malware-corrupted system component would not be able to gain access to the algorithm or data.

After the model runs, the developer receives a performance report on the values of the algorithm's performance along with a summary of the data characteristics. Finally, the algorithm owner may request that an encrypted artifact containing information about validation results is stored for regulatory compliance purposes. Then, the data and the algorithm are wiped from the system. "As an innovator and as an algorithm developer, I now have access to a large world of data that would have taken me years and cost millions of dollars to accumulate—if I ever could. I don't have to worry any longer that the IP [intellectual property] that I've struggled to develop is at risk for being exploited any longer. I now have access to the data I need to develop and validate my algorithms, and I know I can do that in a safe way. That's a much better world to be in from a healthcare and technology perspective than where we are now.", says Blum.[9]



How BeeKeeperAI^{TM} works

BeeKeeperAI^{TM} uses novel technology and standardized workflows

STEP 1
An algorithm owner submits their encrypted algorithm up to the BeeKeeperAI™ platform

STEP 2
The algorithm gets wrapped in a secure computing container

STEP 3
The algorithm is brought to the data owner's environment. The encrypted data it's going to run against is enclosed in a secure encrypted environment*

STEP 4
The data and algorithm are brought together. The algorithm runs and returns the values

STEP 5
The only thing that leaves the secure computing enclave is a confidential performance report and the general characteristics of the data set that it ran against

STEP 6
Finally, the data and algorithm are destroyed.

*Third parties validate that data security has met all compliance requirements

Benefitting the entire healthcare AI ecosystem

BeeKeeperAI will enable developers to access the data they need without creating privacy or security risks for data stewards. “Bringing together these technologies creates an unprecedented opportunity to accelerate AI deployment in real-world settings,” says Dr. Rachael Callcut, MD, CDHI Director of Data Science.[10]

Through BeeKeeperAI, data stewards with the mission of advancing scientific innovation will gain the ability to collaborate with each other along with algorithm owners while maintaining their commitment to privacy and security.

The solution will make it easier for innovators to create AI algorithms that benefit more people in more places and deploy the AI to providers and patients faster and at a lower cost. Whether it’s battling the next pandemic, diagnosing the disease earlier, or enabling more personalized medicine, patients will ultimately be the most important beneficiaries of this technological leap.

Learn more about Azure confidential computing and Intel SGX.

Learn more about Fortanix Confidential Computing Manager.

---

[1] The impact of artificial intelligence in medicine on the future role of the physician, NCBI, 2019

[2] Prognostication of patients with COVID-19 using artificial intelligence based on chest x-rays and clinical data: a retrospective study, The Lancet, 2021

[3] Using AI to predict breast cancer and personalize care, MIT, 2019

[4] Artificial Intelligence That Reads Chest X-Rays Is Approved by FDA, UCSF, 2020

[5] The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database, Nature, 2020

[6] The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database, Nature, 2020

[7] UCSF Joins Forces with Tech Companies to Eliminate Data-Sharing Risks, HealthLeaders Media

[8] Microsoft Azure Forms Collaboration to Enhance AI in Healthcare, Hit Infrastructure

[9] UCSF Joins Forces with Tech Companies to Eliminate Data-Sharing Risks, HealthLeaders Media

[10] Microsoft Azure Forms Collaboration to Enhance AI in Healthcare, Hit Infrastructure

Big Data Artificial Intelligence Machine Learning