Intel & Enzolytics Whitepaper:

Optimizing Empathetic AI to Help Healthcare Providers Cure Deadly Diseases

The Package

The Premise of P4 Medicine

> *"The global pandemic changed the world, forcing governments and organizations to re-examine how we operate, especially in technology."*
>
> Michelle Stout
> Senior Director
> Global Business Acceleration
> Governments, Markets & Trade
> Intel Corporation

The process of healthcare is evolving from the "reactive" (doctor takes part when disease appears) to a more "anticipatory" role as exemplified by P4 Medicine. As a concept coined by Biologist Leroy Hood,[1] P4 Medicine is:

- Predictive: Our genetic makeup can predict the diseases we are at risk for.
- Preventative: Medicine that refocuses care on the future rather than the present.
- Personalized: Every human is genetically different and unique. This kind of approach will give autonomy to individuals as they transition from health to disease.
- Participatory: This pillar is a crucial aspect that involves first convincing physicians of the potential of adopting newer technology. They can then educate patients about the opportunities, challenges, and responsibilities of adopting the technology to change the entire medical community's mindset.

The premise of P4 Medicine embraces a targeted approach to manage a person's overall health, as opposed to simply managing the disease. The future of medicine is individualized care through P4 Medicine.

Intel Corporation

- Clifton Roberts
  Director, Governments, Markets, & Trade
- Nikhil Deshpande
  Senior Director, Data Platforms Group

Enzolytics, Inc.

- Charles Cotropia
  Chief Executive Officer
- Dr. Gaurav Chandra
  Chief Operating Officer
- Dr. Joseph Cotropia
  Chief Scientific Officer

Denver Scientific

- Purvi Patel
  Co-founder, Chief Executive Officer & Chief Scientific Officer

[1] Mauricio Flores, Gustavo Glusman, Kristin Brogaard, Nathan D Price, and Leroy Hood, P4 Medicine: How Systems Medicine Will Transform The Healthcare Sector And Society, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4204402/, US National Library of Medicine

The Promise

Intel recognizes the premise of P4 Medicine, committed to fulfilling the promise that Artificial Intelligence (A.I.) delivers to Healthcare. P4 Medicine and A.I. share characteristics that have, and continue to, challenge today's thought leaders and solution architects:

- P4 Medicine aims to be predictive; A.I. is programmed to be prognostic
- P4 Medicine seeks preventative measures; A.I. is designed to be anticipatory
- P4 Medicine embraces personalized care; A.I. learns by being adaptive
- P4 Medicine encourages physician/patient participation; A.I. thrives on inclusive and participatory acts

Human advances have been driven by the evolution of tools, machines, and discoveries that enlarge our regular abilities. However, in the past, there has been minimal development in the adaptation of technology to understand the human emotional mind – the part that controls our empathy. At Intel, we firmly believe that A.I. can change that.

Structuring human driven A.I. interactions, improved to create confided relationships between AI and people, displays the most significant opportunity for human progression in this cutting-edge era. Empathetic A.I. outsources healthcare tasks and decisions to rational machines, freeing up time for healthcare professionals to engage in empathetic care that cultivates trusting relationships with their patients."[2] Intel has made incredible achievements in this arena, including in:[3]

- Cervical Cancer Screening – Intel developed an algorithm with MobileODT that identified cervix types based on images. This approach helps to prevent ineffectual treatment, assisting healthcare providers give proper referrals.
- Radiology Improvement – C.T. image specialists and Intel used Intel 's Deep Learning Deployment Toolkit and Intel Xeon processors. The result: medical imaging solutions that use deep learning to aid clinicians in assessing results and diagnosing illness.
- Precision Medicine – AI-based solution running on Intel Xeon processors deployed in hospitals for precision medicine applications for more individualized treatment.
- Tumor Detection – Working with large medical images where deep learning is used to train models to help detect tumors using Intel Xeon processor-based systems.
- Genomics Research – Intel and Synthetic Genomics, Incorporated (SGI) used deep learning to gain new insights on protein sequences tagging.
- Skin Cancer Detection – Physicians used the Intel® Movidius™ Neural Compute Stick (NCS) to help screenings for patients reducing wait times.
- Brain Imaging Analysis – Intel and the Princeton Neuroscience Institute optimized software to monitor brain activity to better decode human thought, helping advance neurocognitive therapy for mental illness.

The Puzzle

While Empathetic A.I. is promising, scientific and healthcare-related studies generate large amounts of invaluable data. Additionally, notable characteristics of healthcare-related data present unique, but all too familiar challenges:

[2] Angeliki Kerasidou, Empathy, Compassion, and Trust Balancing Artificial Intelligence in Healthcare, Bulletin of the World Health Organization, https://www.who.int/bulletin/online_first/BLT.19.237198.pdf?ua=1

[3] Intel Corporation, A.I. for Social Good, https://www.intel.com/content/www/us/en/artificial-intelligence/ai4socialgood.html

- Data Resides in Multiple Locations – From different source systems to various departments, healthcare-related data is harvested in other formats (e.g., text, numbers, analog, digital, images, video, etc.). And while departments like radiology produce data in the form of images, there is an increased variety of data collection mechanisms and creation.[4] This increase includes an onslaught of patient-generated data harvested from fitness monitors, blood pressure sensors, and contact tracking devices.
- Structured and Unstructured Data – While efforts to standardize data capture through Electronic Medical Record software have been valiant, unstructured data capture has been reluctantly tolerated through the years to accommodate electronic medical record end-users in favor of patient-centered care delivery. Hence, aggregation and analysis of healthcare-related data often pose difficulties when preparing data models.
- The Breakneck Velocity of Breakthrough Discoveries – Throughout history, human beings have demonstrated an unyielding curiosity about how the human body works. The paradox is that as this curiosity continually enhances our understanding, we simultaneously develop breakthroughs shattering our previously held knowledge.
- Complexity – Gone are the days of finite patient information. In a digital world, there is a need for healthcare providers, scientists, and researchers to connect very complex individual systems that store information about the most complex individual system known to humankind: the human body. Thus, data management of such complex systems, often captured in disparate platforms, requires a sophisticated technology yet to be perfected, amplified, and adopted.
- Evolving Regulatory Environments – Regulatory, legal, and compliance pressures have forced Healthcare and life sciences practitioners to do more with less, to compete with these pressures while coping with new and emerging challenges. For instance, healthcare providers are constantly challenged by governing actions focused on data privacy protections designed to mitigate risks associated with the improper collection, handling, storage, and use of patient and subject-matter data. Consequently, regulatory, legal, and compliance risk management programs should be equally modernized to meet evolving healthcare laws and regulations.

The Path

Traditional approaches to data management and consequent A.I. analysis will not continue to be optimal for healthcare research and development. A revolutionary approach is required; an approach that can handle numerous sources, fearlessly confront unstructured data in all of its inconsistencies, invariabilities, and complexities; all within an ever-evolving regulatory environment. Understanding these data management challenges in terms of relationships among data is paramount. Expressly, the more intricate the relationship is understood and appreciated, both within and among data sources, the more comprehensive the problem's portrait becomes. Graph analytics has emerged as a novel methodology to examine structures of large networks of data; networks of data produced from variant data sources and concluding observed patterns.

"While artificial intelligence, 5G technology, quantum computing, and other innovations are projected to have enormous impacts on the relationship between machines and people, another novel technology will forever change how we analyze big data and gather insights: traversing and scaling knowledge graphs."

Joanne Kim
Duke University

Inspired and encouraged by the DARPA HIVE Project, Intel responded to "build a graph analytics processor that can process streaming graphs thousands of times faster, and at much lower power than current processing technology."[5]

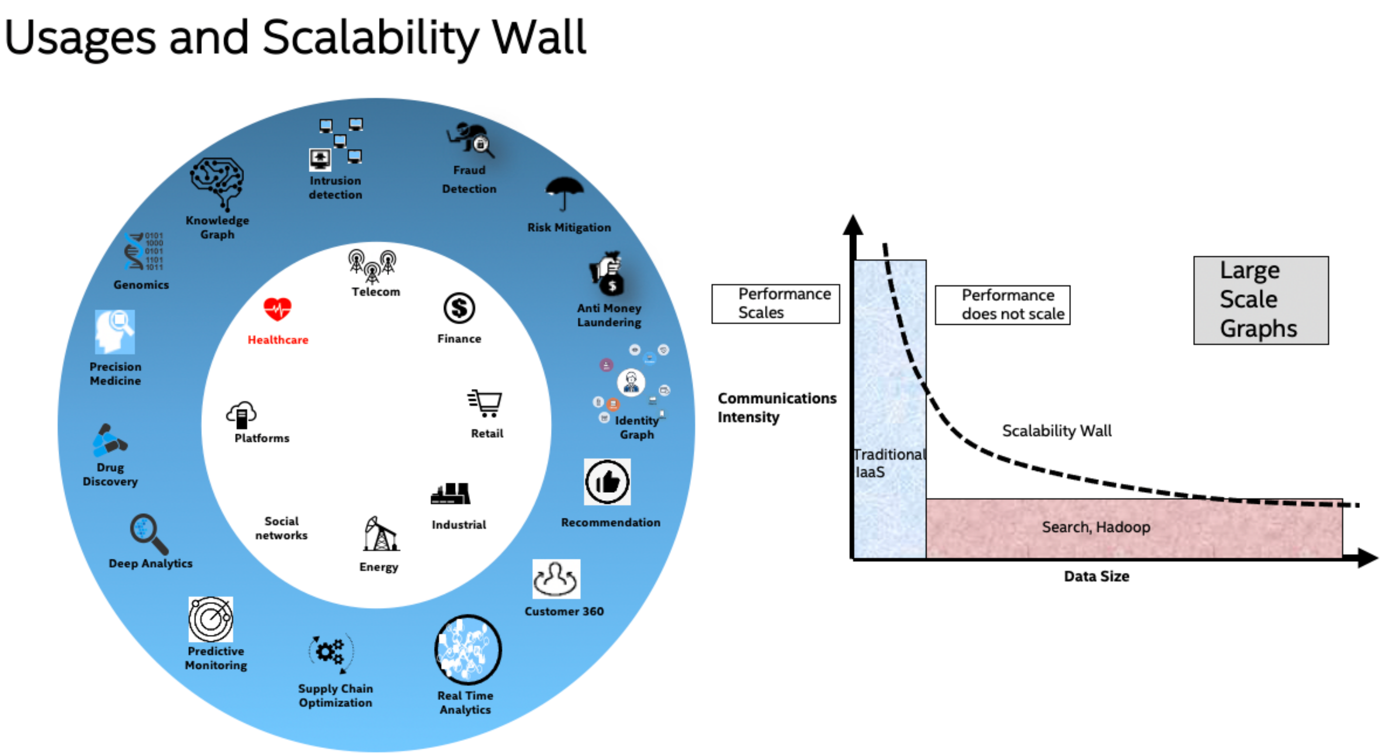
Intel long ago recognized graph analytics as the next frontier for big data and thus sought to create an engineering solution that could drive near real-time actionable insights, informing life-changing decisions by

[4] David Hoffman, Riccardo Masucci, Intel's AI Privacy Whitepaper, https://www.intel.com/content/www/us/en/artificial-intelligence/solutions/ai-privacy-policy-white-paper.html
[5] Dr. Bryan Jacobs, Hierarchical Identify Verify Exploit (HIVE), Defense Advanced Research Projects Agency, https://www.darpa.mil/program/hierarchical-identify-verify-exploit

public and private sector actors. With the data deluge, representing large-scale unstructured data poses a significant challenge, as traditional structured data analytics techniques do not lend well to this type of data. Graph representation of data enables a schema of independent analytics that scales with the data size and data types. Knowledge Graphs – a significant class of graphical representations of data – are expected to play a substantial role in several spaces: investment insights, fraud prevention, cybersecurity, government analytics, enterprise A.I. analytics, social media insights, drug discovery & repurposing, and predictions of virus mutations.

However, traditional compute and systems architectures are not designed to traverse large graphs rapidly due to the "Scalability Wall." In some sense, the efficiencies such as caching that traditional architectures take advantage of become the primary sources of ineffectiveness when traversing graphs. Systems need to be designed with different balances among compute, memory, and network when the objective is to capture rapid insights from a large-scale graph.

Under the Hood: Intel PIUMA Technology



As demonstrated, heavy workloads with communications, and less so with compute, do not optimally scale as these workloads are confronted by more compute. The fundamental issue to solve here is not the compute, but rather the communications bandwidth.

Harnessing innovation, Intel designed and developed Intel® Programmable Integrated Unified Memory Architecture (PIUMA). PIUMA is Intel's new hardware architecture model with an ability to capture real-time insights as it traverses through graphs with blazing fast speed compared to legacy hardware architectures.

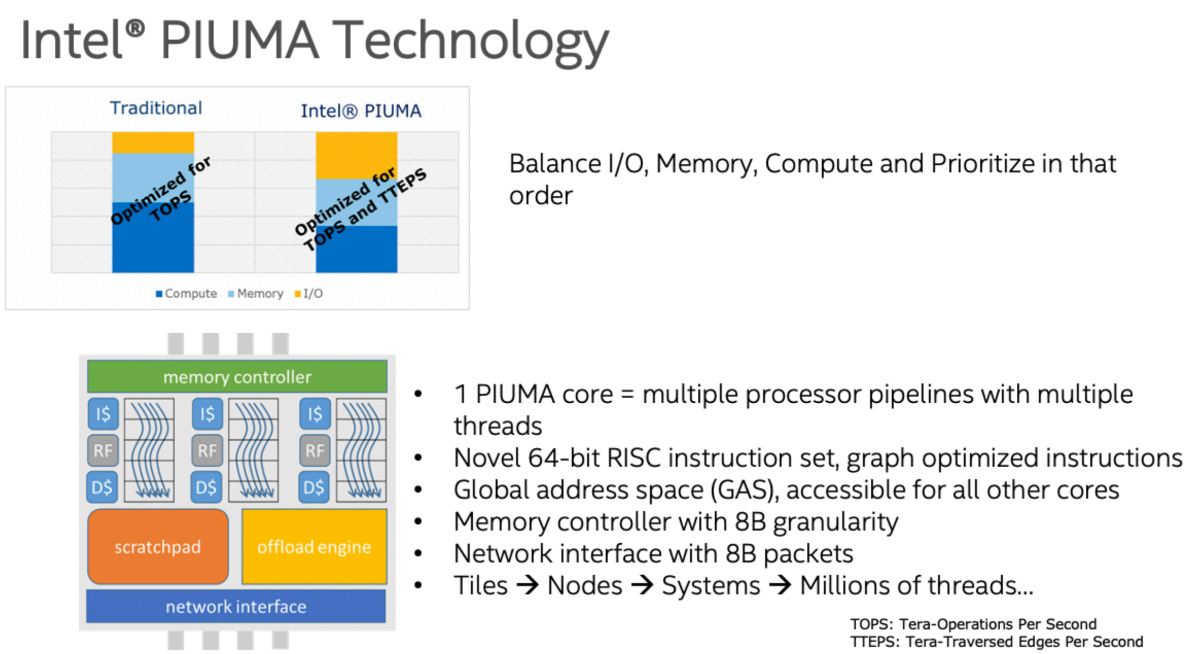
Intel® PIUMA enjoys no switches, extremely high bandwidth, low latency, a balanced compute to I/O ratio, and the capability to process random accesses to account for unique patterns of graphs. Intel® PIUMA was meticulously designed to drive extremely low latency analytics on hundreds of Terabytes and Petabytes of scale data.

> *"Intel® PIUMA and graph analytics will add velocity to drug discoveries, repurposing of existing drugs, and predictions of virus mutations."*

Some of the innovations include:

- multiple processor pipelines, with multiple threads
- novel 64-bit RISC instruction set, graph optimized instructions
- Global Address Space (GAS); and,
- memory controller with 8-64B granularity Network interfaces with 8B packets.

With these innovations, PIUMA enables a balanced architecture across compute, memory, and I/O, as illustrated here:

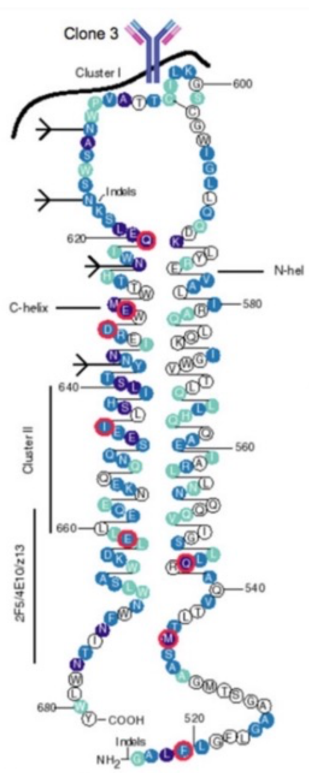


Intel has successfully architected a hardware solution that has forever transformed real-time data analytics, with enormous potential to revolutionize the digital economy. Its many applications and numerous positive implications will make Intel® PIUMA an impactful technology solution that can work in conjunction with various software architectures to help drive better decision-making. Further, Intel®'s PIUMA will indeed affect business operations and society, as it fills the gap and needs for quick, deep insights that ultimately add meaning to big data.

The Paradigm

Case Study: Enzolytics Uses Technology to Create Fully Human Monoclonal Antibodies in Confronting Viruses

Dr. Joseph Cotropia discovered a human cell line producing a specific human antibody that targets a genetic amino acid sequence designated KLIC on HIV's outer envelope's surface. Discovering this sequence is like finding a "needle in a haystack "and retrieving the needle. The immutable sites on the virus must be identified, and the fully human monoclonal antibodies that target those sites must be created and characterized. The amino acid sequence for the monoclonal (Clone 3) antibody has been determined and named by Dr. Cotropia. This antibody can lock onto the KLIC epitope. Once bound by the neutralizing antibody, the virus cannot infect a human cell and cannot ultimately reproduce. Directed against this highly conserved epitope on the HIV gp41 transmembrane protein, the monoclonal Clone 3 Antibody has been epitope-mapped and characterized. Patented[6] by Dr. Cotropia,[7] this HIV monoclonal antibody has been successfully tested in five international labs. Clone 3 antibody neutralized [at an IC90[8]] 95% of all HIV primary isolate strains (41/43) —across all clades and groups — against which it was tested.



HIV gp41 viral stalk

To date, there are almost 6000 different strains of HIV-1 now listed in the Los Alamos National Laboratory HIV Database.[9]

For an antibody to be effective, it has to attack a neutralizable site on the virus that is always present and does not mutate. Simply put, where a site on a virus is highly conserved and immutable from strain to strain. Knowing the binding site for the monoclonal Clone 3 antibody on the HIV-1 virus and then examining the Coronavirus amino acid sequence, a correlation in the structures has been identified by Dr. Joseph Cotropia and Dr. Gaurav Chandra.

A KLIC Epitope is a conserved target on the viral envelope that makes the virus vulnerable to the antibody's neutralizing action

An infinite number of distinct anti-HIV and anti-Coronavirus monoclonal antibodies exists – some disease neutralizing, some perhaps of no benefit, and some that may be disease enhancing. Thus, specific antibodies that broadly neutralize are necessary to provide effective therapy and prevention of disease. These applications in using a monoclonal antibody in passive immunotherapy are accomplished by using Enzolytics' proprietary method of producing broadly neutralizing monoclonal antibodies. However, a primary focus on identifying immutable binding sites on the virus and then creating monoclonal antibodies that bind to such immutable sites is required will universally and durably neutralize the SARS-CoV-2 virus. In this way, targeting highly conserved epitope sites—the virus cannot mutate around the therapy and thereby escape effective passive immunotherapy protective and preventive treatment.

[6] NIH, National Library of Medicine, https://pubmed.ncbi.nlm.nih.gov/?term=Cotropia, pubmed.gov
[7] About Dr. Joseph Cotropia, http://enzolytics.com/about-enzolytics-inc/
[8] IC90 is the concentration that will inhibit 90% of all HIV virions
[9] Los Alamos National Laboratory, HIV Molecular Immunology Database, https://www.hiv.lanl.gov/content/immunology/maps/ab/gp160.html

A.I. in Identifying /Immutable Target Sites on HIV and SARS-Cov-2

Dr. Gaurav Chandra recognized A.I.'s potential and is a firm believer in the idea that "The eyes see what the mind knows." He believes that A.I.'s most significant potential lies in the innovator's ability to harness its power and improve upon established goals and inventions. Based on knowledge of these homologous viral structures, targeting the corresponding "Achilles Heel" site on a virus, an expected conserved immutable and neutralizable site was identified. Consequently, Dr. Chandra worked with Denver Scientific's genetics[10]molecular biology data science team to screen more than 50,512 Coronavirus and 87,336 HIV isolates, the largest known repository of HIV and Covid-19 isolates in the world. From this very complex and extensive A.I. analysis, conserved sites immutable on HIV and Covid-19 were identified.

> *"The situation at the beginning of the first years of the AIDS epidemic was that my care culminated into sitting at the bedside of these human beings, holding their hands as they died"*
>
> Dr. Joseph Cotropia
> Enzolytics, Inc.

The use of AI helped to confirm the sequence determined throughout two decades of Dr. Cotropia's career. Additionally, another seven conserved sequences have also been identified using AI technology, while in another breakthrough, 19 conserved sequences over the entirety of the 50,512 Coronavirus isolates analyzed have been identified. These immutable sites on the SARS-CoV-2 virus have also been confirmed as existing (100%) in the U.S. SARS-CoV-2 virus variants that have surfaced in United Kingdom, Brazil, and South Africa. The Center for Disease Control (CDC) recently reported these Coronavirus variants as "variants of concern." These variants show "evidence of an increase in transmissibility, increased hospitalizations or deaths, a significant reduction in neutralization by antibodies generated during previous infection or vaccination, reduced effectiveness of treatments or vaccines, or failures in diagnostic detection." This report exemplifies the findings by Doctors, Cotropia and Chandra.

A.I. will allow the examination of the virus's numerous different strains to identify other conserved sites and produce additional monoclonal antibodies targeting them. The goal is to create a "collection" or "cocktail" of antibodies for therapeutic use. We recognize that there are now over 16,000 different known variations or strains of the Coronavirus, each slightly different due to mutation.

> *"Our Intelligence is what makes us human, and AI is an extension of that quality."*
>
> Yann LeCun
> Author
> Artificial Intelligence Revealed

[10] https://denverscientificai.net/

The Prescription

Public Policy to Fuel a Promising Trifecta

As a trifecta, Intel technology, P4 Medicine, and the efforts of Enzoyltics requires effective public policy; policies that will fuel this trifecta as a promising thought leadership experiment; a necessary journey of experimentation as humanity bravely marches towards cures for society's most deadly diseases; towards a more Empathetic A.I. Cultivating a technology-neutral public policy, regulatory, and government investment environment is paramount to this end. Hence, there are five primary policy drivers to cultivate trust, empathy, and velocity that will help to materialize the potential described herein:

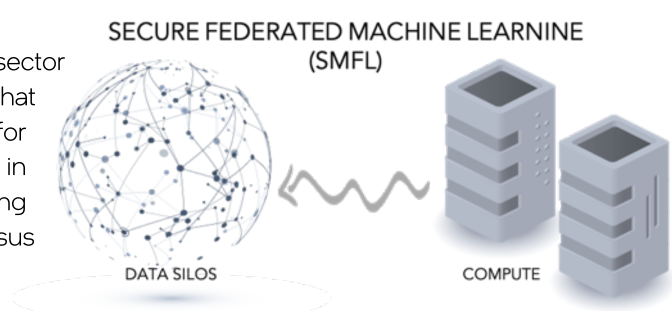
1. Perpetuate Public-Private Partnerships

The social significance of the work described herein makes it essential to establish strategic partnerships guided by global health directives through commitments and strategic alignments among governments, the scientific community, and private and philanthropic organizations. Specifically, a progressively symbiotic relationship between the public and private sector actors is needed to succeed in their mutual goals of confronting society's most deadly diseases. What we see in the use case of Enzyolytics is the use of data and AI to underscore the viability of its innovations. In the case of the U.S., the government purchases more than half of the healthcare in the country. Public sector solutions established to reduce cost and improve care need be accelerated to meet the urgent demands we face in confronting history's most deadly and ecomonimcally impactful diseases society is facing.

Enzolytic's approach exemplifies how Artificial Intelligence can assess millions of virus sequences and identify the conserved segments essential for virus survival. We can then harness Artificial Intelligence's power, coupled with our advanced techniques in immunotherapeutics, to create universal, durable, broadly effective treatment targeting all virus variants.

2. Foster Secure Federated Machine Learning (SFML)

A significant challenge that healthcare innovators like Enzolytics face in its work to harvest data is that data is stored across numerous source systems, within various healthcare-related departments, and is harvested in many formats, resulting in data silos. While an optimal solution would be to consolidate and house these oceans of data within a single location, doing so would be impractical, demanding the consumption of vast amounts of resources to satisfy such a requirement.



To confront these challenges head-on, public and private sector actors should partner to develop and publish case studies that demonstrate best-known-methods. Additionally, tool kits for employing Federated Learning practices should be employed in the research of deadly diseases. SFML brings processing mechanisms to the data source for training and inferencing versus requiring that agencies migrate data to one place.[11] Such data federation ensures privacy and security of both data and machine learning models, offering additional confidence in post-deployment insights. SFML ensures that (i) data remains

[11] Nikhil Deshpande, Clifton Roberts, Reimagining Public Data to Advance Ai: Maximizing Insights from Federal Data, https://blogs.intel.com/policy/2019/08/16/re-imagining-public-data/#gs.7eud5k

in place and compute moves to the data, and (ii) both compute and data are protected at the hardware level from confidentiality and data integrity attacks. Such assurances are especially important in the area of healthcare involving the most intimate of personally identifiable patient information.

Numerous SFML studies have illustrated its effectiveness, allowing for fast deployment & testing of smarter models, low latency, and less power consumption. SFML employs a combination of privacy-by-design techniques to ensure data de-identification, data protection, and insights security. As the commercial sector plans to consume federal data for algorithmic expression(s), the business interest of those participants must be protected through security techniques ingrained at the lowest level of hardware – silicon!

3 Standards and Frameworks Implementation to Address Structured and Unstructured Data Challenges

We stated earlier that aggregation and analysis of healthcare-related data often pose difficulties when preparing data models. Specifically, data represents a complex field, since data is subject to a variety of regulatory regimes, which include privacy, data sovereignty, localization, and cross-border transfers. This type of subject-matter variance cultivates disharmony in common regulatory approaches and stifles the development of regulations that rely on voluntary standards in evolving technical requirements for AI data use cases.

Access to large, reliable datasets is essential to the development and deployment of robust and trusted AI. Standards and guidelines play an important role in developing approaches for access to AI datasets. However, they need to be carefully defined based on different use case contexts and consider common ethical concerns, including privacy regulations. Data-related standards, such as metadata and format interoperability standards, can:

- make available public sources of information in structured and accessible databases
- create reliable datasets (while employing data de-identification techniques) for use by all AI developers to test automated solutions and benchmark program(s) quality; and,
- foster incentives for data sharing between the public and private sector and among healthcare industry players like Enzolytics and Denver Scientific.

In similar areas highly affected by privacy regulations, international standards have been successful in defining useful mechanisms to support a harmonized approach to regulatory objectives. Examples of such standards include the Do Not Track standard (under W3C) and anonymous signatures and authentication standards (under ISO/IEC JTC 1 SC 27)

4 Ethical Liberation of Datasets for Evolving Regulatory Environments

One growing technology trend is the increase in mechanisms for data collection and creation. Personal data is not just collected from individuals who provide it for particular uses, but also observed and gathered by sensors in connected devices and derived or created through further automated processing. In fact, the percentage of data coming directly from individuals is decreasing compared to the information that is collected in our increasingly connected society and inferred through machine learning technologies.

While ethical data processing is built on privacy, the increasing amount of data collected, processed and inferred using encryption and de-identification (full anonymization) techniques are indeed capable of protecting patients' privacy. Nonetheless, achieving de-identification will require increasingly complex practices because re-identification will be increasingly possible in deep learning-driven environments. Differential privacy (DP) techniques have emerged in the last years as viable solutions to minimize privacy risks, adding "noise" to scramble personal data. Furthermore, in the academic

and research community, Homomorphic encryption (HE) seems particularly promising as it allows computation on encrypted data, enabling AI tasks without the need to transfer personal information.
Government investment in following practices is in U.S. national interest to further drive innovations in the marketplace:

- The development of international standards for algorithmic explainability
- Promotion of diversity in datasets
- Research & development
- The transparency of outcome-based studies about de-identification techniques like DP and HE[12]

Finally, in practice, data de-identification techniques should be uniformly applied across agencies to ensure that (i) data is anonymized and (ii) combining anonymized data across silos does not result into the re-identification of subjects.

5 Empathetic A.I. Requires Minimization of Bias

The paradox of data is that while data reflects societal issues, societal issues can influence data collected, possessing inherent racial, gender, economic, regional, and other types of biases. To date, nearly 200 human biases[13] have been defined and classified, any one of which can affect how we make decisions. As such, there is a probability that the consumption, processing, and sharing of such data will lead to discriminatory insights that under-represent the macrocosm, which could erode trust between humans and machines that learn. Such bias could particularly be amplified in SFML where learning occurs on the data node rather than the combined dataset.

Constructing datasets while optimizing AI algorithms to avoid bias are interdependent functions. Continued investment by government in fostering private sector actors' access to large and reliable datasets is essential to the development and deployment of innovative, trusted, inclusive, and empathetic AI, especially in concerted efforts to confront deadly disease. Specifically, greater diversity in datasets will reduce the risk of unintended bias. Diversity in the teams working to curate and then release datasets for AI can also address training bias. A diverse, inclusive team of individuals with different skills and approaches to the curation and release of datasets ensures more holistic and ethical designs of AI algorithms. Teams at Intel Corporation, Enzolytics Inc., and Denver Scientific exemplify this public policy approach.

The Pinnacle

In closing, Enzolytic's use case underscores a novel approach of how Artificial Intelligence can assess millions of virus sequences and identify the conserved segments essential for virus survival. Through effective public and private sector symbiosis, as well as the fostering of a technology-neutral regulatory environment, the genuine power of A.I., coupled with advanced techniques in proprietary technologies illustrated in the Enzolytics immunotherapeutics, will no doubt create universal, durable, and broadly effective treatment targeting all virus variants. This will allow medical practitioners the much-needed time to focus on the intangibles of the medical profession: to show their patients empathy, to hold their hands through tough times, and to assure them there is hope in overcoming their conditions.

[12] David Hoffman, Riccardo Masucci, Intel's AI Privacy Whitepaper, https://www.intel.com/content/www/us/en/artificial-intelligence/solutions/ai-privacy-policy-white-paper.html
[13] Cognitive Bias Codec, https://www.visualcapitalist.com/wp-content/uploads/2017/09/cognitive-bias-infographic.html

"As governments develop and implement horizontal AI Strategies, it will be important to consider specific plans for AI in healthcare."

Mario Romao
Global Director, Health Care Policy
Intel Corporation