SOLUTION BRIEF

Data Analytics & AI
Intel® Optimized Software

Simpler Development, Faster Performance for Enterprise Deep Learning Solutions

Arya.ai

The power of artificial intelligence (AI) is beginning to impact some of today's most data-driven business sectors. Insurance and banking companies, for example, are beginning to use deep learning applications to better understand and predict customer behavior based on text, call logs, social media activity, transactional data, and more. Humans and traditional rules-based applications can be overwhelmed by such large amounts of diverse data. Deep learning applications thrive on it and can continuously learn from historical results and from new data flows to make increasingly accurate decisions.

Arya.ai provides a relatively simple and cost-effective pathway for financial services organizations and other enterprise customers to integrate AI into their business models. VEGA*, the company's deep learning workbench, simplifies the development and deployment of deep learning solutions. Arya.ai also offers modular, deep learnings apps that can be integrated into existing workflows for claims processing, underwriting, fraud detection, lending, check processing, and more. Several Fortune 500 companies are already working with Arya.ai to jump-start their AI journeys.[1]

VEGA—A Deep Learning Workbench for the Enterprise

VEGA simplifies the design, deployment and management of deep learning applications. From drag-and-drop neural network design, to automated tuning and one-click cloud deployment, VEGA streamlines AI development. It is designed for both beginners and experienced data scientists, and functions as both a development environment and a centralized AI platform that can scale as companies grow their AI footprint.

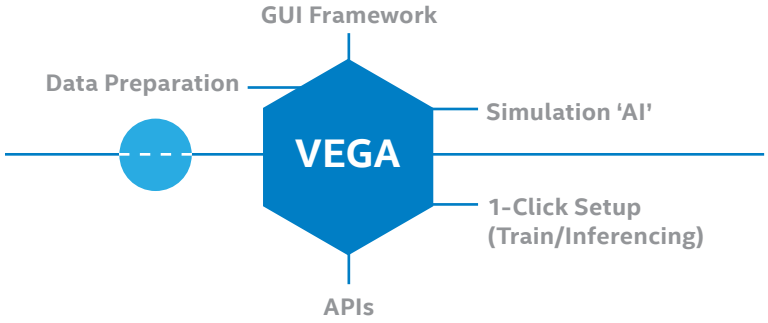

Figure 1. VEGA provides a complete AI workbench that simplifies development, deployment, and management for enterprise customers.

INTEL®
AI BUILDERS
MEMBER

Enterprise-Ready Modules for Financial Services Organizations

Potential usage models for deep learning are nearly limitless. Almost any decision-making process that relies on large amounts of data is likely to be a good candidate. Most businesses have many such use cases and using VEGA to develop in-house expertise may pay big dividends down the road.

Arya.ai also offers pre-built modules for key functions in the financial services industry. These complex AI solutions fuse multiple deep learning models to address broader business goals.

- **Automated Underwriting and Pricing** evaluates insurance applications, determines risk, and generates personalized offers and premiums based on each applicant's profile. It combines deep personalization with fast throughput.
- **Customer 360** looks at a wide range of structured and unstructured customer data to provide a more complete and nuanced view of each individual. It can help to identify prospective customers and to create customized offers based on a better understanding of opportunities, risks, and requirements.
- **Lending AI** optimizes collections to improve an organization's bottom line. It identifies customers who are at risk of defaulting and helps to identify effective interventions. It also flags potentially fraudulent activity.
- **Automated Check Processing** replaces or complements manual and rules-based methods. The self-learning neural networks can read both handwritten and machine-printed data on checks and can authenticate signatures with a high level of confidence.
- **Credit Card Fraud Detection** provides real-time monitoring with autonomous learning that can detect both simple and sophisticated attacks and can adapt quickly to changing patterns of fraudulent activity.

For more information and options, visit Arya.ai.com

AI in Action—Insurance Claims Processing

Insurance claims processing provides a good example of how Arya.ai modules can reduce costs and improve accuracy for complex business decisions. Most insurance companies rely on a combination of rules-based applications and expert humans to evaluate the eligibility of claims and determine how much money, if any, should be awarded.

There is typically a lot of data to sift through for each claim, data that may include text, images, calls, videos, agent feedback, transactional data and more. It's a slow, laborious, and costly process that introduces significant overhead for the insurer and frustrating delays for customers. Yet accuracy cannot be sacrificed without increasing risk, since fraud is an ever-present concern. The move to online solutions has simplified many aspects of claims processing, yet the volume and diversity of relevant and available data have also skyrocketed.

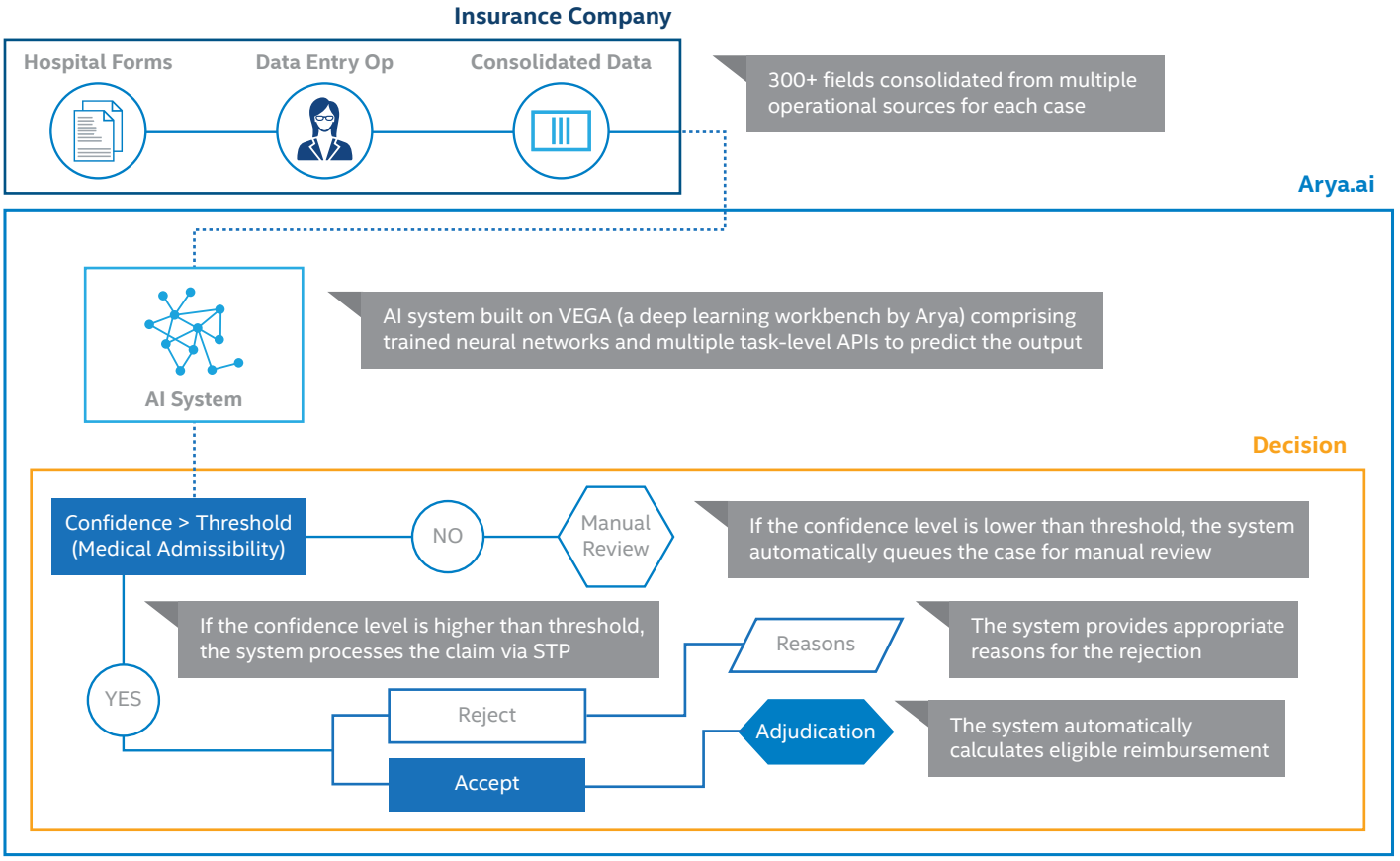

Figure 2. ClaimsNet* uses multiple, fused neural networks to improve and accelerate insurance claims processing.

Arya.ai offers its Automated Claims Processing module to help companies streamline claims processing while improving results. This module provides an AI-based solution for pre-authorizing and adjudicating new claims, and for identifying potentially fraudulent transactions. The heart of the solution is ClaimsNet*, a fused collection of neural networks. Unlike more traditional machine learning strategies based on complicated rules, ClaimsNet learns to process claims from historical data. There is no need for complex modeling and manual programming.

For each claim, ClaimsNet analyses more than 300 fields of consolidated data from multiple operational sources. It determines a confidence level for each claim. If the confidence level is lower than a specified threshold, the claim is automatically queued for manual review. If the confidence level is sufficiently high, the claim is either accepted, and the eligible reimbursement amount is automatically calculated, or the claim is rejected, in which case the reasons for rejecting it are provided along with the decision.

Learning loops are built into ClaimsNet, so the neural networks continue to learn from successes and failures. New data sources can also be integrated as they become available, and built-in APIs simplify integration with existing workflows. The ability to learn and improve without manual intervention makes ClaimsNet a powerful starting point for integrating AI into existing environments.

Optimized Software Boosts Performance

As AI is deployed into mainstream business environments, cost-effective performance becomes increasingly important. Specialized hardware accelerators add cost and complexity to data center solutions and make it more challenging to deploy and scale AI solutions. To answer this challenge, Arya.ai turned to optimized software packages from Intel that help to improve the performance of AI workloads on Intel® architecture.

Arya.ai recently integrated Intel® Distribution for Python* and Intel® Optimization for TensorFlow* into its VEGA platform. Python has become one of the most popular languages for machine learning and deep learning applications, and TensorFlow is one of the most popular deep learning frameworks.

Intel Distribution for Python and Intel Optimization for TensorFlow use the Intel® Math Kernel Library (Intel® MKL), which provides optimized algorithms for the basic mathematical operations that underpin many scientific, technical, analytic, and financial applications. An important component of Intel MKL is the Intel® Math Kernel Library for Deep Neural Networks (Intel® MKL-DNN). This specialized library provides optimized algorithms for many of the most popular and compute-intensive operations that are used in neural network training and inference solutions.

The algorithms in Intel MKL are optimized to efficiently use the execution resources in the latest Intel® Xeon® Scalable processors. Workloads are distributed across multiple cores and threads, and memory and cache usage are optimized to improve data flow to the cores. Intel MKL also supports Intel® Advanced Vector Extensions 512 (Intel® AVX 512), which enables single instructions to be processed simultaneously on multiple data elements stored in 512-bit vector registers. Optimizing software for this technology is known as vectorization and can dramatically increase performance for neural network processing, and for other operations that can be parallelized in this way.

5.95X Faster Inference Performance[1]

To verify the performance benefits for their deep learning solutions, Arya.ai engineers ran two versions of ClaimsNet on a two-socket server configured with Intel® Xeon® Platinum 8153 processors:

- **A baseline version** using the open source distributions of Python and TensorFlow
- **An optimized version** using Intel Distribution for Python and Intel Optimization for TensorFlow

During the performance tests, anonymized health insurance claims and related data were processed with ClaimsNet to provide claim handling recommendations. The claims were sorted into three groups, including:

- **Accepted claims** with recommended award amounts
- **Rejected claims** with the reasons for rejection
- **Claims queued for manual intervention** because they could not be assessed with sufficiently high confidence

The performance for the optimized version of ClaimsNet was 5.95X[1] that of the unoptimized version (Figure 3). This performance boost can help organizations process more claims at higher speeds on the same infrastructure footprint. It can also help them scale the solution more easily as workloads increase and new data sources are integrated.

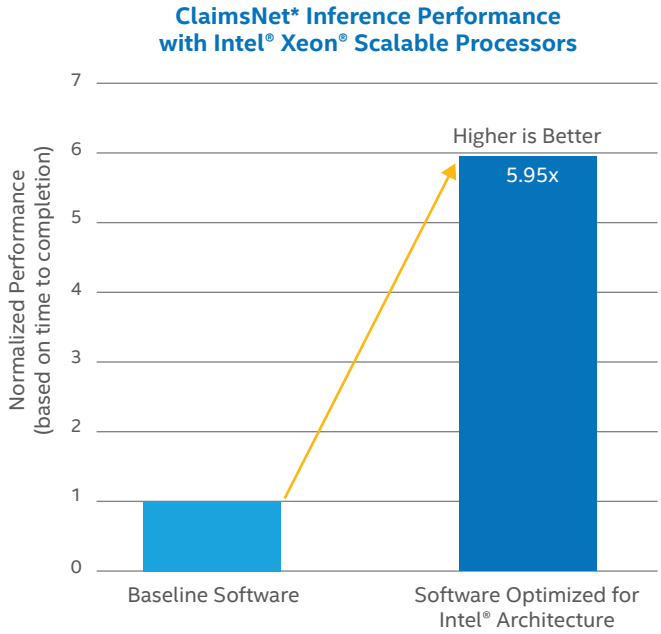

Figure 3. Replacing the standard distributions of Python* and TensorFlow* with Intel® Distribution for Python and Intel Optimization for TensorFlow boosted ClaimsNet performance by 5.95X[1].

Performance results are based on testing as of 19 June, 2018 and may not reflect all publicly available security updates. See configuration disclosure for details. No product or component can be absolutely secure.

Better Solutions for the Broader AI Community

Intel works with open source and commercial software developers to optimize a wide range of popular frameworks and applications for each new Intel® processor generation. In the case of open source software, such as Python and TensorFlow, Intel typically up-streams the optimized code into the main distribution of the application. With this approach, users can obtain highly optimized performance out of the box on Intel Xeon Scalable processors, as well as many other Intel processors.

This software optimization work is helping to fuel the AI revolution by allowing companies such as Arya.ai to run demanding deep learning workloads on industry-standard servers, without the need for costly, specialized hardware platforms. Implementation is simpler, and AI applications can be integrated more easily with other enterprise workloads.

Conclusion

Arya.ai and other members of Intel® AI Builders are building AI solutions that are beginning to transform certain data-intensive industries, such as financial services, manufacturing, and retail. With ClaimsNet and other enterprise-ready deep learning apps, Arya.ai is providing a quick onramp for organizations looking to improve speed, efficiency, and accuracy for critical functions in the insurance, banking, and investment industries. With VEGA, they are making it easier for organizations in all industries to bring the power of deep learning to bear on virtually any data-dependent decision-making process.

We believe these and other emerging deep learning solutions mark the beginning of a broad move toward mainstream AI adoption that will drive ever-deeper improvements in efficiency, automation, responsiveness and personalization. As this transformation unfolds, Intel® hardware and software technologies will continue to play a foundational role, helping vendors and customers to speed innovation, while reducing cost and complexity.

Arya.ai is a member of the Intel® AI Builders Program, an ecosystem of industry leading independent software vendors (ISVs), system integrators (SIs), original equipment manufacturers (OEMs), and enterprise end users who have a shared mission to accelerate the adoption of artificial intelligence across Intel® platforms.

To learn more, visit Intel AI Builders at: **https://builders.intel.com/ai**

To learn more about Arya.ai, visit: **https://arya.ai/**

[1] Tests performed by Arya.ai as of 19 June, 2018 to compare performance using an inference workload for a typical health insurance use case using a batch load of 100. System Configuration: Two-socket server configured with 2 x Intel® Xeon® Platinum 8153 processor (2.00GHz, 64 cores), 384 GB DDR4@2666MHz memory (12 x 32 GB DIMMS), 129 TB Intel® SSD (SC2KB48), CentOS* Linux release 7.4.1708 (Core) (3.10.0-693.el7.x86_64). Baseline Software: Standard open source Python*, standard open source TensorFlow*. Optimized Software: Intel® Distribution for Python* 3.6, Intel® Optimization for TensorFlow 1.6.0.

Performance results are based on testing as of 19 June, 2018 and may not reflect all publicly available security updates. See configuration disclosure for details. No product or component can be absolutely secure.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more information go to www.intel.com/benchmarks.

Intel does not control or audit third-party benchmark data or the web sites referenced in this document. You should visit the referenced web site and confirm whether referenced data are accurate. Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No computer system can be absolutely secure. Cost optimization scenarios described are intended as examples of how a given Intel- based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Intel does not guarantee any costs or cost reduction. Check with your system manufacturer or retailer or learn more at intel.com.

Optimization Notice: Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice.

No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

Intel disclaims all express and implied warranties, including without limitation, the implied warranties of merchantability, fitness for a particular purpose, and non-infringement, as well as any warranty arising from course of performance, course of dealing, or usage in trade.

This document contains information on products, services and/or processes in development. All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest forecast, schedule, specifications and roadmaps.

The products and services described may contain defects or errors known as errata which may cause deviations from published specifications. Current characterized errata are available on request.

Copies of documents which have an order number and are referenced in this document may be obtained by calling 1-800-548-4725 or by visiting www.intel.com/design/literature.htm.

Intel, the Intel logo, and Xeon are trademarks of Intel Corporation in the U.S. and/or other countries.

*Other names and brands may be claimed as the property of others.

© 2019 Intel Corporation. 0219/RA/MESH/PDF 338802-001US