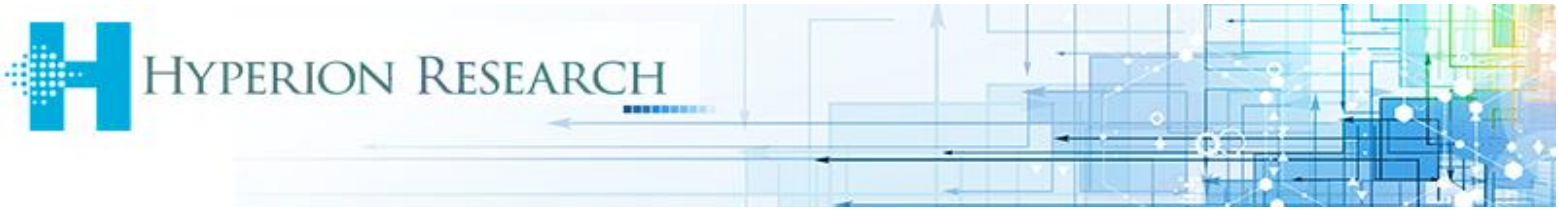

Taxonomy

Hyperion Research Taxonomy of HPC-Enabled Artificial Intelligence Use Cases

Steve Conway, Earl Joseph, Alex Norton, Melissa Riddle, and Mark Nossokoff
May 2021

HYPERION RESEARCH OPINION

Artificial intelligence (AI) is in the early stages of transforming science, engineering, and important aspects of everyday life. This document presents a taxonomy of AI use cases that are in various stages of development. What they all share is reliance, to a greater or lesser extent, on high performance computing (HPC). This taxonomy doesn't attempt to cover the entire AI market; it focuses instead on use cases at the forefront of AI research and development, where HPC support is often indispensable. Because these use cases help to define where the mainstream AI market will be heading in the future, this taxonomy may also be of interest to people outside of the global HPC community.

This AI taxonomy expands on the shorter list of use cases closely tracked as today's most economically important HPC-enabled, repetitive AI activities: precision medicine, automated (and semi-automated) driving systems, fraud and anomaly detection, affinity marketing, business intelligence, cyber security, and IoT/edge computing/smart cities. The expanded list is based on AI global research studies and reviews of the relevant literature. While some of the added use cases don't represent large economic opportunities yet, each promises to become more important in the near future. The expanded list adds a fair number of use cases but cannot claim to be exhaustive, and there is overlap among the categories. Some, such as cyber security and fraud and anomaly detection, are horizontal applications that are important for many of the listed vertical use cases.

Historically, HPC has advanced the state of the art in science and engineering through the most powerful computing resources contemporarily available. As third-party clouds and enterprise data centers are exploiting AI in advanced data analytics and business intelligence, they now require powerful computing and data movement resources traditionally exclusive to HPC data centers. Users in many areas are now turning to HPC to support the AI methodologies of machine learning, deep learning, and graph analytics.

Note: AI includes machine learning, deep learning, and other technologies.

Note: This page is intentionally blank.

TABLE OF CONTENTS

SITUATION OVERVIEW

AI use cases, enabled today by HPC technology, are often at the forefront of R&D. HPC vendors and users should find these definitions useful as an overview of workload types in this young, fast-growing area of the worldwide HPC market. Individuals and organizations active in the mainstream AI market may find the definitions useful as indicators of the mainstream market's future direction and for considering whether to add HPC technology and expertise to their computing resources.

DEFINITIONS

Advanced Scientific and Engineering Research

HPC-supported AI methodologies are already being used in every major scientific and engineering research field, including the broad domains of physics and chemistry, to help achieve advances in theoretical and applied knowledge. AI benefits in scientific and engineering research vary widely from greatly reducing the need for modeling and simulation (e.g., by eliminating irrelevant areas of the problem space) to sifting rapidly through massive volumes of literature or source data.

Automated (and Semi-Automated) Driving

For decades now, nearly all vehicles have been designed with the help of HPC server systems—using methods including finite element modeling and analysis, mechanical computer-aided engineering, civil engineering, structural analysis, computational fluid dynamics (CFD), crash, NVH (noise, vibration and harshness), and solid modeling. AI methods, especially machine and deep learning, have expanded this methodological toolkit to help meet the challenges of designing vehicles with growing abilities to operate autonomously—especially by "rightsizing" simulation demand in parametric modeling and data processing in finished designs.

- Conceptual Design: an early design stage in which the overall form and uses of the vehicle are laid out with the help of drawings, small-scale models and preliminary computer simulations.
- Component Design: determining the forms and functions of parts of vehicles as needed by the vehicle as a whole. Component design may be performed by the vehicle designer or by a contractor.
- Full Vehicle Design: the complex, iterative process (parametric modeling) of designing an operational vehicle by optimizing the interaction of many variables.
- Testing and Standards: design work to meet safety and other regulatory requirements in the automotive industry. RAND Corp. has estimated 8.8 billion miles of physical testing would be needed to attain 95% consumer trust in fully automated vehicles, and this would take 400 years. Applying HPC simulation and AI methods should reduce this time frame to 5-10 years.
- Insurance Underwriting: already part of HPC use in financial services, this task evaluates risks an insurer takes in covering vehicles, so that the insurer can decide which risks to take and how much to charge for each one. HPC's speed has been especially useful for real-time telephone rate quotes. AI promises to accelerate underwriting substantially.
- Accident Reconstruction: determining how an accident happened in order to assign liability for insurance purposes. Many nations and states are in the process of developing liability regulations for automated vehicles, especially to decide whether drivers or vehicles (i.e., carmakers) are at fault in specific situations.

- **Cyber Security:** fully automated vehicles will handle large volumes of data in highly networked environments that attract hackers. HPC-enabled AI is already starting to be used to design data security features for these vehicles.

Aerospace

The aerospace applications include start to finish design of aircraft, as well as simulation and development of components within aircraft. AI has become critical in accelerating simulations for designing new aircraft, including the use of AI to uncover some new solution spaces for different designs. AI is also heavily used in the cyber security application space, which is a critical component in the aerospace sector.

- **Conceptual Design:** an early design stage in which the overall form and uses of the aircraft are laid out with the help of drawings, small-scale models, and preliminary computer simulations.
- **Testing and Standards:** design work to meet safety and other regulatory requirements in the aerospace industry.
- **Accident Reconstruction:** determining how an accident happened in order to assign liability for insurance purposes and address any design problems to help prevent similar accidents in the future.
- **Cyber Security:** modern commercial aircraft already employ substantial automation and handle large volumes of data in highly networked environments that may attract hackers. HPC-enabled AI will increasingly be used to design data security features for these aircraft.
- **Equipment Placement, Crew Assignment, and Route Optimization:** the airline industry has been seeking the magic formula for this computationally complex use case for several decades. The goal is to ideally coordinate the locations of aircraft and available crews, while also optimizing routes to minimize fuel and labor costs—an instance of the proverbial "traveling salesman" problem. HPC-enabled AI can help, and quantum computing advanced by HPC users will also be useful for this task.
- **Revenue Optimization:** another complex computational task that involves optimizing aircraft seating configurations and pricing; times of day, week and year; aircraft types (e.g., regional jets for non-long-haul routes); and other factors.
- **Automated Flight Control:** a use case important for drones, historically in the military and increasingly under consideration by private-sector companies (e.g., for package delivery).

Bio-Life Sciences

This sector includes various application spaces related to biology, pharmacology, medicine, and genomics. AI is being implemented in accelerating drug discovery, as well as the biggest use case so far, in developing AI-based models to help diagnose and treat medical ailments. Precision medicine promises to be a high-profit AI use case, and an area of heavy investment today.

- **Precision Medicine:** a medical model based on customizing healthcare, with medical decisions, practices, and products tailored to the individual patient rather than on a set of standardized procedures applicable to all patients. HPC is helping to lead the way toward precision medicine, especially through DNA sequencing and acting as a decision-support tool for physicians and other providers.
- **Genomics:** a branch of biotechnology concerned with applying the techniques of genetics and molecular biology to the mapping and DNA sequencing of sets of genes or the complete genomes of selected organisms, with organizing the results in databases, and with applications of the data (as in medicine or biology). HPC clusters are often used to process the

results of purpose-built gene sequencing machines. AI support is particularly useful for data-intensive applications, such as comparative genomics.

- **Proteomics:** a branch of biotechnology concerned with applying the techniques of molecular biology, biochemistry, and genetics to analyzing the structure, function, and interactions of the proteins produced by the genes of a particular cell, tissue, or organism, along with organizing the information in databases and with applications of the data. HPC systems are often used to process the associated large data sets and AI accelerates the processing, such as by predicting which protein structures will exhibit desired behaviors.
- **Drug Discovery**: in the fields of medicine, biotechnology, and pharmacology, drug discovery is the process by which new candidate medications are discovered. Historically, drugs were discovered through identifying the active ingredient from traditional remedies or by serendipitous discovery. HPC systems are used to quickly test large (often millions) of drug candidates (small molecules) for lock-and-key fit with disease agents.
- **Drug Formulation and Testing:** HPC-enabled AI can also help determine the optimal dosage and delivery format for a new drug (e.g., pill, liquid, capsule, injection), as well as supporting the data-intensive needs of multi-phase clinical trials to accelerate regulatory approval and time-to-market.
- **Bioinformatics:** the application of computer technology to the management of biological information. Computers are used to gather, store, analyze, and integrate biological and genetic information which can then be applied to gene-based drug discovery and development.
- **Agricultural Research:** the application of computer technology to the science of optimizing agricultural production and quality.
- **Epidemiology/Public Health:** epidemiology is the study of the etiology and spread of disease outbreaks that affect, or threaten to affect, large segments of a population. Public health is a general term for the study of phenomena that affect the health of a human population.
- **Radiology:** HPC-enabled AI is increasingly used to read diagnostic medical images with greater-than-human accuracy and to optimize radiation dosing.

Cyber Security

Many large and small organizations take advantage of HPC and AI for cyber security needs, including threats from internal and external members of the organization, as well as many corporations in handling the fraud and anomaly detection. AI techniques have become critical in tracking and uncovering further fraud and cyber threats, training systems to learn patterns and expose potential issues quicker than humans.

- **Fraud and Anomaly Detection:** this "horizontal" workload segment centers around identifying harmful or potentially harmful patterns and their causes by detecting "faint signals" with AI methods such as graph analysis, semantic analysis, and machine and deep learning. Fraud is wrongful or illegal deception for financial or personal gain. Cyber fraud may include the exploitation or illegitimate annotation of data.
- **Cyberattack:** an attempt by an outside party to gain unauthorized access to stored data or disrupt services, usually via network intrusion (hacking). Cyber attackers have breached electrical grids, planting potentially harmful software in the process.
- **Insider Threat:** the unauthorized appropriation or exploitation of data by a person with access to the organization owning the data. Espionage is one insider threat category in government.

Insider theft is also a serious problem in the private sector, or may point to cybersecurity crime, significant errors, or other anomalies that may deserve further investigation.

- **Anti-Terrorism and Law Enforcement:** this AI use case focuses on identifying planned or completed criminal activity. The graph analytics method can be especially useful because of its ability to home in on suspicious activity via relationships among individuals, locations, and behaviors. In one notable instance, HPC-enabled AI was used to identify a terrorist sleeper cell by reviewing 250,000 cell phone accounts and one million related connections in order to identify one account of interest, that of the sleeper cell's leader.

Defense

This workload centers around applications such as surveillance and signal processing; encryption; command, control, communications, and intelligence (C3I); geospatial image management and analysis; defense research; weapons design; and other national security applications. In addition, national security organizations are fielding applications that work to identify and track potential security threats through database-oriented pattern-matching applications. These applications will be run in conjunction with traditional security applications such as cryptography and image analysis.

- **National defense:** this workload centers around applications such as surveillance and signal processing; encryption; command, control, communications, and intelligence (C3I); geospatial image management and analysis; defense research; weapons design; and other national security applications. In addition, national security organizations are fielding applications that work to identify and track potential security threats through database-oriented pattern-matching applications. Although these applications may not always be numerically intensive, they will be developed and used by organizations that are firmly rooted in technical computing markets. These applications will be run in conjunction with traditional security applications such as cryptography and image analysis.
- **Military Planning/War Games:** HPC-enabled AI will increasingly be used for the data-intensive simulation of potential military engagements.
- **Battlefield Operations:** HPC-enabled AI will also be useful for wide-area situation analysis and command-and-control related to live battlefield operations.
- **Weapons Development and Testing:** HPC modeling and simulation has been used for decades to help develop advanced military weapons, including materials such as ceramics designed to withstand projectile attacks. AI in this context plays a role similar to its function in designing commercial vehicles.

Energy

These applications revolve mainly around the modeling and simulation associated with major oil and gas companies, conducting seismic modeling and reservoir simulation to uncover current and new oil reservoirs. Applications in this space also include power grid simulations and various new energy source simulations and deployments.

- **Energy Data Science:** an application used by management of energy firms and related government agencies for planning energy resources and predicting price fluctuations.
- **Exploration and Production (E&P):** major oil and gas companies vie with each other to acquire and operate very large supercomputers. In general, the more powerful the computer, the farther it will allow the companies to "see" high-potential sites for drilling that may be miles below the land or sea surface. HPC and HPC-enabled AI can also optimize the energy production process.

- **Maintenance Prediction for Field Equipment:** O&G field equipment is expensive and subject to hard use. Using HPC-enabled AI to predict maintenance needs can save valuable labor and opportunity costs.
- **Operational Support for Well Sites:** HPC-enabled AI can also be useful for monitoring, modeling, and managing well site operations.
- **Smart Power Grids:** AI promises to react quickly to local power grid breaches to prevent cascading failures that could otherwise result in regional blackouts.
- **Alternative Energy Research:** HPC is already used heavily to support research into renewable energy sources. AI can accelerate this research.

Economics/Financial

This workload centers on applications such as econometric modeling, portfolio management, stock market and economic forecasting, and financial analysis. The segment includes both trader and computationally intensive non-trader tasks. HPC has been used for this workload category since the 1980s. HPC-enabled AI promises to complete these tasks with greater speed and accuracy.

- **Portfolio Optimization:** the process of choosing the proportions of various assets to be held in a portfolio in such a way as to make the portfolio better than any other arrangement, according to some criterion. The criterion will combine, directly or indirectly, considerations of the expected value of the portfolio's rate of return and the return's dispersion and possibly other measures of financial risk.
- **Pricing Exotic Instruments:** a derivative that is more complex than commonly traded "vanilla" products. This complexity usually relates to determination of payoff. The category may also include derivatives with a non-standard subject matter (i.e., underlying) developed for a particular client or a particular market.
- **Global Risk Management:** management of the worldwide financial risk incurred at any point in time by a company that trades stock, bonds or other financial instruments, based on calculating the aggregate risk incurred by all of the firm's traders.

Government Laboratories and Research Centers

This workload centers on government-funded research and development institutions. These organizations are generally funded at a national or multinational level and may combine both purely scientific research with research in areas of national priority (e.g., cancer research) and/or research for defense-related programs. These users are less bound by strict economic constraints than those performing applications in product development environments. These centers don't normally offer degree programs for students.

Humanities and Social Sciences

In the past decade, HPC has increasingly been used to address complex tasks outside of the "hard sciences." All of these use cases are data-intensive enough to benefit substantially from AI methods.

- **Archeology:** archaeology researchers at the UK's University of Bradford are using their HPC system to revolutionize data collation, effectively preserving endangered or destroyed archeological heritage sites across the world. On an even grander scale, so-called galactic archeologists are employing HPC to determine how, when, and why various galactic bodies formed or disappeared.
- **Cultural Anthropology:** anthropologists have used HPC to help study the Maya peoples of Central America and how cultures around the world have used technology.

- **Historical Linguistics:** linguists at the University of Reading (UK), Ohio State University and other institutions employ HPC to help illuminate the historical evolution of languages and language families.

Internet of Things, Edge Computing and Smart Cities

These three uses for HPC-enabled AI are closely related to each other. The vast majority of "things" on the Internet of Things (IoT) are edge devices, and smart cities can be viewed at dense IoT nodes.

- **Internet of Things:** ordinary and non-ordinary devices can now be connected to each other and a central hub via the internet. These devices, which can be as small as sensors and as large as cars, planes, boats, etc., can all share information with each other. China's national IoT plan extends out to the year 2030 and is based on an HPC architecture. Other nations will also make use of HPC and HPC-enabled AI to support IoT networks and their massive data processing requirements.
- **Edge Computing:** a relatively new approach for distributed computing where some or all of the necessary computation is done directly at or near data sources, such as vehicles and traffic sensors, medical devices, product manufacturing lines, military sites, and other data-generating locations. In many cases, the limited computing power available at or near the data sources is adequate. In cases where deeper analysis is required, a small subset of the results from each local edge computing location may be sent to a data center or cloud for processing on a high-performance computer. In these cases, the "small data" from each source location is typically combined in the data center or cloud to produce "big data" for the more powerful computer to process. This variety of edge computing is important when the goal is to obtain situational awareness and decision support for issues affecting larger areas, such as citywide traffic control, epidemic and severe storm tracking or military operations.
- **Smart City Operations:** in 2010, China launched what was then the world's most powerful supercomputer, named Tianhe-1. The city and province of Guangzhou (formerly Canton), with a population of more than 14 million people, funded 60% of this $330 million behemoth and dedicated most of its use to traffic congestion and other smart city functions. Other urban areas around the world are pursuing smart city plans as necessities for future economic competitiveness. Frequent functions that will benefit from HPC-enabled AI include plans for construction, vehicle traffic, energy use, air quality and others.

Marketing and Sales

In a recent Hyperion Research study, 36% of the respondents said they use HPC systems for live business operations, as opposed to HPC's predominant historical role in support of upstream R&D. Much of the newer commercial activity focuses on sales and marketing activities.

- **Sales Analysis and Revenue Optimization:** historically, large companies would ask their enterprise computers relatively simply questions, such as, "Who were our top 10 salespeople in France during the last quarter?" Today, in order to remain competitive, these firms need to go much deeper, with questions such as, "Who were our top 10 salespeople in France during the last quarter, and why? Was it related to their territories? Their managers? The products they carried? Their personal abilities?" Companies use the results of these complex analyses to make changes designed to boost revenue in future quarters.
- **Affinity Marketing:** this segment covers the use of HPC and HPC-enabled AI to promote products or services, typically using complex algorithms to discern potential customers' demographics, buying preferences and habits. A stretch goal of some companies is to model

entire shopping malls, including the purchasing preferences of all individuals entering the mall for potential real-time promotion of store inventories.

Weather Forecasting

This workload centers on applications such as atmospheric modeling, meteorology, weather forecasting, and climate modeling. This segment includes systems dedicated to these tasks primarily in the government and defense segments.

- Climate Research: the use of HPC to study the dynamics and effects of past and ongoing climate change, and to predict future change. AI is increasingly used to sift through huge data volumes involved in this research field and to augment traditional simulation findings with analytics-based algorithms.

FUTURE OUTLOOK

In a recent worldwide study, 89% of HPC user sites said they now employ one or more AI methodologies. The worldwide market for AI-focused HPC server systems will expand more than four-fold by 2024 to reach $3.9 billion, up from $918 million in 2019 (32.9% CAGR). HPC-enabled AI will find many standalone new uses, but for the most part, AI methodologies will be applied to HPC workloads in combination with long-standing modeling and simulation methods such as computational fluid dynamics and structural analysis.

This taxonomy will be updated regularly to reflect the changing nature of the marketplace. Both vendors and users are encouraged and welcomed to give any suggestions or recommendations to help better define and size the market.

About Hyperion Research, LLC

Hyperion Research provides data-driven research, analysis and recommendations for technologies, applications, and markets in high performance computing and emerging technology areas to help organizations worldwide make effective decisions and seize growth opportunities. Research includes market sizing and forecasting, share tracking, segmentation, technology, and related trend analysis, and both user & vendor analysis for multi-user technical server technology used for HPC and HPDA (high performance data analysis). Hyperion Research provides thought leadership and practical guidance for users, vendors and other members of the HPC community by focusing on key market and technology trends across government, industry, commerce, and academia.

Headquarters

365 Summit Avenue
St. Paul, MN 55102
USA
612.812.5798
www.HyperionResearch.com and www.hpcuserforum.com
