

Is Artificial Intelligence (A.I.) Ready to Run a Factory?

Craig Eric Seidelson

University of Indianapolis, USA, seidelsonc@uindy.edu

Abstract: With smart factory investment expected to increase 20% year-on-year over the next five years and total investment expected to reach \$275 billion worldwide by 2027, the use of Artificial Intelligence (A.I.) to manage operations is receiving considerable attention. This paper takes an in depth look at how factory data is being generated, stored, processed, transferred, trained and ultimately validated using A.I. The conclusion is that deep machine learning is more than capable of controlling devices. Yet, research shows only 14% of smart manufactures would describe their A.I. efforts as successful. The problems are cost and application. Smart manufacturing is almost exclusively done by multi-billion dollar operations. Is this money well spent? Factories aren't closed, linear systems. In these chaotic systems infinitesimal changes in any one of the myriad of input variables are capable of producing disproportionate changes in output values. As a result, no matter how much scrap, downtime, sales or on-time delivery data a company collects actual values will diverge exponentially from what existing A.I. algorithms are predicting. Until more research is done predicting dynamic, nonlinear systems A.I. will not be capable of running a factory without human involvement.

Keywords: Artificial Intelligence, Forecasting, ERP, Machine learning, Chaos

Introduction

The concept of machine-to-machine (M2M) communication isn't new. In the early 1970's Theodore Paraskevakos patented the transmission of "information from a calling telephone to a called telephone" (Link Labs, 2015). By the late 1990's, M2M communication described devices sharing data over networks. Over the following decades more and more data would be shared over a global Internet. Today, the IoT (Internet of Things) encompasses some 50 billion data sharing devices worldwide (Techjury, 2021). It's estimated by 2025 the number of IoT devices will reach 75 billion (Telecom Review, 2020).

Data is being produced on an unprecedented scale. For example, in 2016 the world generated 2,500,000,000 GB of data each day. Nearly 90% of all data that ever existed prior to 2016 came into existence during 2016 (Mabkhot, 2018). As impressive as the ability to generate data has become it pales in comparison to the ability to store it. Worldwide data storage capacity (i.e. GlobalStorage Sphere) doubles every four years. Today, the GlobalStorage Sphere stands at 6,200,000,000,000 GB (Reinsel, 2020). Roughly 60% of this space is currently being utilized (Techtarget, 2018).

Cyber-Physical Systems

The ability to generate and store continuous streams of data has allowed engineers to construct "virtual factories." In this digital world "discrete event modeling" is used to describe the flow of products through production steps. "Agent-based modeling" allows programmers to place production elements (i.e. people, facilities, products, orders, etc.) inside simulated environments to observe system behaviors (Anylogic, 2020).

Virtual factories are routinely used by the Volvo Group Global to validate proposed production changes before they're introduced into actual plants (Jain, 2014). However, simulating factories has a number of limitations. A factory is an open system. Any number of outside variables (i.e. absenteeism, training, product mix, on-time delivery, inventory levels, machine breakdowns, scrap, etc.) impact what's happening inside the system. Factories, like any open system, never settle down in to a steady state. This presents serious problems for modelling. Cyber-physical systems (CPS) address this weakness by continuously feeding information back into models. The result is computations monitor and control physical processes while feedback loops allow physical processes to update computations (UC Berkeley EECS Dept., 2019).

The ability for machines to collect, share, analyze and act upon vast amounts of data requires extremely fast and flexible computer processors (CPUs). Multicore CPUs are capable of executing billions of calculations per second (GHz). CPS, however, requires more - clusters of multicore processors working in parallel. One such

cluster, the IBM AC922, is made up of 4,608 computer servers. In one second this supercomputer can perform 200 quadrillion (e.g. 200 with 15 zeros) calculations (Bryner, 2018). Fast processor speeds and memory transfers are at the heart of smart manufacturing.

Artificial Intelligence (A.I.)

A factory's enterprise resource planning (ERP) systems transfer all of the accounting, production, supply chain, sales, marketing and human resources data using in-memory, relational databases. ERP provider SAP requires at a minimum 60 GB of storage capacity. The amount of data storage needed to run a business might seem like a lot – but it isn't considering sensors, solenoids and actuators on a single IoT enabled device will typically generate 5GB of data per week. Even with clustered processors it can take on the order of hours to extract structured (and unstructured) data in multiple disparate formats; transform it into a format that can be analyzed; and then load it into a data warehouse. Collectively, delays in data extraction, transformation and loading is referred to as ETL lag. Machine learning in near real time at factories must compensate for ETL lag. This is commonly done using a lambda (λ) architecture which gives quick answers based on some of the data and accurate answers based on all of the data.

The continuous flow of high variety data at high volume into λ architecture is accomplished by breaking data up into manageable chunks using a queuing system “like Kafka and a streaming system like Storm, Spark or Flink” (Zweben, 2016). Algorithms on each coding layer become inputs to other algorithms on other coding layers. In this way batches data transfers between layers once an hour or once a day (Huilgol, 2017). As more-and-more data passes through more-and-more computations on more-and-more layers “deep learned” model parameters are self-adjusted (i.e. “trained”) to make better predictions about future data. To the extent existing labeled data matches predicted data the model is “validated.” To the extent new data matches predicted data the model is “tested” for use. New and old data continuously “retrain” batch layer programs in the hope of improving how well algorithms can monitor and control physical processes. Periodically, model predicted data is uploaded to serving layers for factory managers to view using NoSQL key value queries.

Because the batching and serving layers are operating on the full data set, these machine learned algorithms are the most accurate. Accuracy, however, comes at a high price. Batch layers needs to “store an immutable, constantly growing master dataset, and compute arbitrary functions on that dataset” (Saxon, 2012). Even with open sourced Hadoop batch clustering systems parallelizing data storage and computations, ETL lag time to propagate new data through batch layers can take hours. Near real-time monitoring and control of physical processes in a factory requires a speed layer.

Machine learned algorithms on a λ speed layer perform computations on the most recent data prior to uploading it to batch layers. Fast reads and writes are possible because speed layer programs, unlike those on batch layers, aren't “continuously re-computing batch views from scratch” (Ulyanov, 2016). “Creating a deep learning model from scratch can take days or weeks to train, because of the large amount of data and rate of learning” (Tan, 2019). Databases on the speed layer, such as Cassandra or HBase, are capable of near real-time monitoring and control of physical processes because they're incrementally updating views (i.e. “transfer learning”) created by analytics programs, such as MapReduce, Hive or Spark, on batch layers.

Incremental programming logic makes the speed layer fast. For example, with clusters of processors operating in parallel speed layer training times decrease from weeks to minutes. Unfortunately, transfer learning and the random database reads and writes mean the speed layer is also, by far, the most complex. The beauty of λ machine learning architecture is system complexity is isolated to computing layers where data only exist temporarily. Once data is uploaded to batch layers it's purged from the speed layer making room for more incoming data and calculations.

In light of data storage, transfer and computations needed to support machine learning, smart manufacturing is primarily done on severless, pay-for-use network clusters known as clouds. Indeed, spending on cloud infrastructure as a service (IaaS) and software as a service (SaaS) reached \$20 billion in 2019 (Sahu, 2020). By 2025, it's estimated that almost half of the world's data will be stored in clouds (Ang, 2020). Per Figure 1, five companies account for nearly 80% of the public cloud (Sahu, 2020).

Public Cloud Market Share 2019

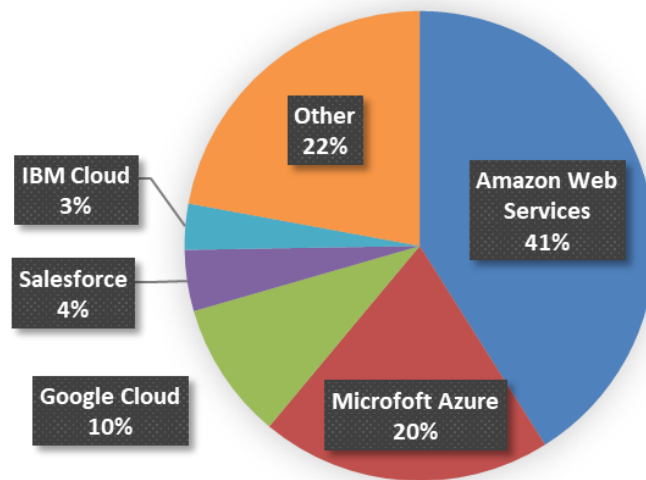


Figure 1. The Public Cloud is Controlled by a Few Providers

With the cloud being an integral part of machine learning it's little wonder that each of the leading public cloud providers offer their own automated machine-learning packages (Knight, 2020). Microsoft has the Machine Learning Studio. Google offers Cloud AutoML and AWS uses SageMaker. Widespread availability of cloud and coding architecture explains in separate studies consultants at McKinsey and PwC found between a quarter to a third of executives plan to roll out A.I. initiatives.

Even with automatically scalable cloud resources and 20x faster 5G downlink speeds capable of supporting 10x more connected devices per unit of floorspace "there's latency concerns when sending data across networks and devices" (Wu, 2020). Near real time machine learning on a speed layer requires bringing data storage and processing off centralized clouds closer to "the edge" where computation outputs are needed. This is particularly true in factories when machine learning data may only be of interest to those applications generating data. By leveraging a wide range of local devices and nearby datacenters, edge computing is key to supporting near real-time CPS (Shaw, 2019).

A.I. Managed Facilities

It has been said that the real manufacturing world is on the verge of "converging with the digital manufacturing world enabling organizations to digitally plan and project the entire lifecycle of products and production facilities" (Hessmann, 2013). Smart factories are an attempt to undertake production without human involvement. Reaching this end involves a "pyramid of four progressing levels: the device level, supervisory control and data acquisition (SCADA) level, manufacturing operations management (MOM) level, and enterprise level" (Industry Week, 2018). A factory is "smart" to the extent answers to the following 10 statements are "yes."

1. Algorithms decide inventory and production levels.
2. Machines provide customers and associates with real time answers to their questions.
3. Machines detect, sort and make corrections for nonconforming products.
4. Algorithms predict quality issues.
5. Algorithms predict maintenance needs.
6. Image recognition locates parts in storage and production.
7. Material handling equipment is self-directed.
8. Algorithms create and validate designs.
9. Production machines are self-operating.
10. Production machines are self-programmed.

Applying the above 10 points yields approximately \$120 billion in worldwide smart factory market

capitalization. It's predicted by the end of 2027 global investment in smart manufacturing will reach \$275 billion (Shah, 2020). It's predicted that smart manufacturing over the next three years will grow at 1.7x the last three years (Columbus, 2019). In spite of these optimistic projections, a 2018 US Census Bureau survey of 583,000 US businesses found only 2.8 percent had actually adopted machine learning (Knight, 2020). A Capgemini Research Institute survey of 1,000 manufacturers with smart factory initiatives underway found only 14% described their deployments as successful (Baggott, 2020). A key issue holding A.I. back is conductivity between devices and analytics. A senior manager responsible for Virtual Methods and IT at a major automobile manufacturer blames poor conductivity on a lack of "products and platforms available, ready to use, that we can simply purchase, implement, and then start using" (Capgemini Research Institute, 2019).

One way around the conductivity problem is piloting AI projects across individual machines, work cells or departments. On the one hand, modular deployments are a good way to realize short term gains. In addition, modules allow users to better understand cost and technical challenges bringing solutions to scale. However, piece meal machine learning isn't always practical or economical. Lambda architecture is inherently complex. Undertaking the implementation and maintenance challenges of keeping batch and speed coding layers in sync may not be worth the effort for individual work centers. Not to mention, algorithm training may not be possible for a small number of machines given the massive amounts of data typically needed. There's also the issue of cost. Cloud providers typically sell storage by the terabyte. This amount of storage is likely more than a few smart machines will ever need. The high cost of coding also needs to be considered. *Element AI* reported "in the entire world, fewer than 10,000 people have the skills necessary to tackle serious artificial intelligence research" (Metz, 2017). It's unlikely that a company with modest A.I. goals has the money to compete with the Silicon Valley for talent. Google's DeepMind A.I. is paying roughly \$300,000 per employee. It follows that "financial issues were the most commonly cited barrier to adoption of smart manufacturing technologies and processes" according to an RTI International report prepared for the National Institute of Standards and Technology (Gallaher, 2016). Of RTI's 80 interview subjects across "a wide variety of smart manufacturing product and service providers, smart manufacturing end-user companies, and industry observers" nearly half (as shown in Figure 2) cited lack of financial resources as holding back A.I.

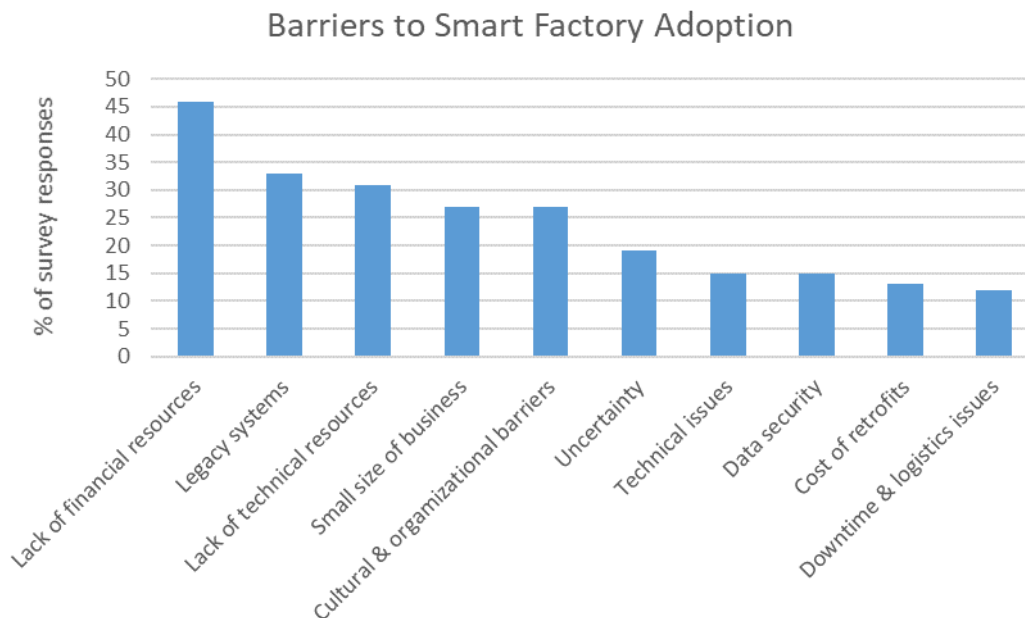


Figure 2. Lack of Financing Slows A.I Adoption in Factories

Given the high cost of smart manufacturing, when it's done it's done almost exclusively by large companies. For example, nine key players in the smart manufacturing value chain (e.g. Siemens AG, General Electric, Rockwell Automation Inc., Schneider Electric, Honeywell International Inc., Emerson Electric Co., and Fanuc Corporation) had a combined market capitalization of \$257 billion in 2019. In the U.S., almost half of all smart factories recorded over \$2.5 billion in sales (Biron, 2017). Wide spread use of A.I. in U.S. manufacturing is unlikely considering companies with over 1,000 employees make up only 0.3 percent of all U.S. factories (U.S. Census Bureau, 2017). The best way to make smart technology more affordable, according to a European Commission report, is through public funding (Digital Transformation Monitor, 2017). The U.S. government is funding smart manufacturing through the Manufacturing USA program. Unfortunately, with federal

contributions typically in the \$70-110 million range at a minimum of 1:1 public to private cost sharing, US government support goes almost exclusively to large companies (Manufacturing.gov, 2020). Researchers at IBM are seeking to make smart technology more affordable by “reduc[ing] the number of bits, or 1s and 0s, needed to represent data—from 16 bits, the current industry standard, to only four” (Hao, 2020). If successful this “could increase the speed and cut the energy costs needed to train deep learning by more than sevenfold” (Hao, 2020).

In addition to cost A.I. use in factories also suffers from lack of compatibility. Every second, 127 devices are connected to the Internet (Gyarmathy, 2020). Unfortunately, many of the 300 IoT platforms supporting these devices use their own “infrastructure, proprietary protocols and interfaces” (Noura, 2019). Lack of interface explains why Materials Resource Planning (MRP), Supervisory Control and Data Acquisition (SCADA) and Enterprise Resource Planning (ERP) systems, all bought at different times, are likely unable to communicate. If, however, A.I. analytics is focused on individual systems with existing machine learning programs (like ERP) issues of cost and compatibility are mitigated.

ERP involves the continuous flow of company data across: finance, quality management, HR, maintenance, procurement, production planning, materials management, sales and logistics. ERP software makes it easy to collect, organize, analyze, and distribute this information because each department is using a single, defined data structure on a common database. The SAP Company, per Figure 3, is the market leader in the ERP business segment.

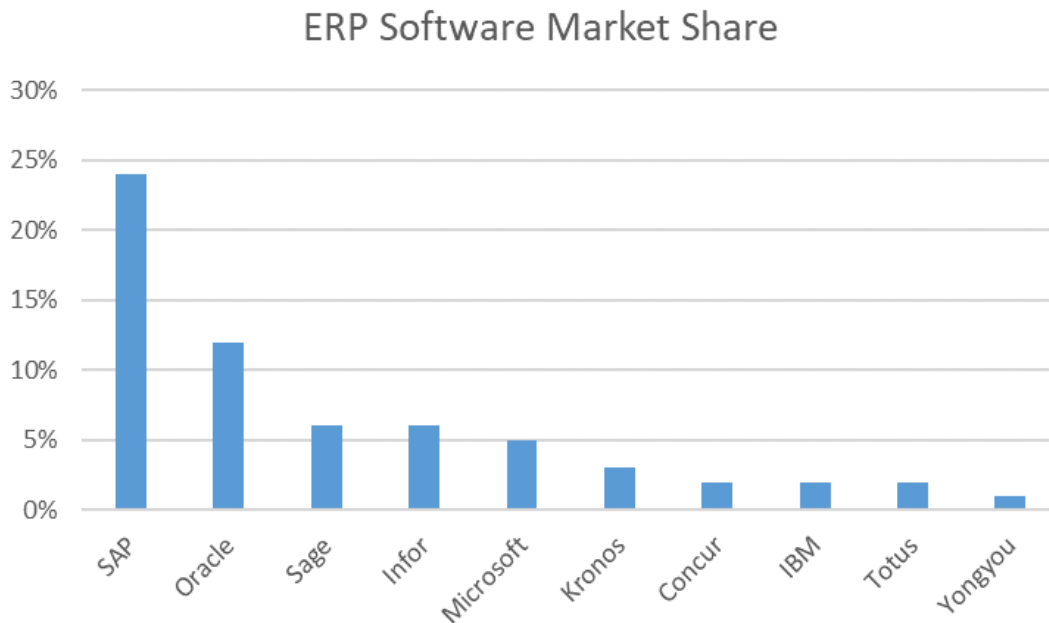


Figure 3. SAP Leads ERP Providers

A.I. in ERP works because standardized, ready-to-use, single vendor end-to-end solutions exist (Infoclutch, 2021). For example, subscribers to the SAP Cloud Platform have no need for onsite data centers. Software-as-a-service and updates are provided through SAP’s private cloud. One of those updates is the SAP Data Intelligence Cloud. Inside the Data Intelligence Cloud unstructured data is sent to the SAP Analytics Cloud for predictive pattern analysis using prebuilt machine learning algorithms. The same happens for structured data on the SAP Data Warehouse Cloud. SAP provided machine learning algorithms are trained on the SAP HANA Cloud. In this arrangement, unlike traditional λ architecture, separation of batch and speed layers to address ETL is not needed. SAP’s multi-cloud landscape has the computing power to bring analytics where the data resides. As a company collects more-and-more data it’s theoretically possible for SAP algorithms to forecast: sales, receivables, payables, scrap, on time deliveries, etc.

SAP Data Intelligence is attempting to predict how company data will evolve over time according to machine learned rules. Predicting future values for linear systems is possible because changes in input values are proportional to changes in outputs. Unfortunately, Business transaction data in a factory is not linear. A very small difference between actual and learned conditions “may have a significant impact at the highest factory

level” (National Research Council, 1995). In other words, small changes in any one of a myriad of factory variables (i.e. number of people, uptime of machines, raw material delivery, throughput rates, inventory levels, process steps, etc.) produce totally unpredictable outcomes. As a result, the evolution of ERP data in a factory fits the definition of mathematical chaos.- i.e. a high dependence on initial conditions whereby value trajectories diverge exponentially over time drastically limiting any possibility of prediction (Michel, 1996). Even though SAP (and other ERP providers) offer A.I packages it’s highly unlikely in these chaotic systems machine learned algorithms will be able to forecast transaction values.

Reservoir computing offers hope for a future when trained machine learning algorithms will be capable of predicting chaotic systems. Reservoir computing ignores the problem of finding solutions to nonlinear equations. Instead reservoir computing algorithms focus on tracking data evolution (Wolchover, 2018). Researchers have been able to extend machine learned prediction timelines of chaotic systems in the lab using new measurements to retrain algorithms “before the trajectories of the reservoir and original systems diverge substantially” (Fan, 2020).

Conclusion

Smart factories are all the rage. The \$4.4bn market in 2019 is expected to continue growing at 20% each year over the next five years. Deep machine learning technology is well established as are the IoT devices, cloud, edge, 5G and coding architecture upon which it depends. This technology, however, isn’t cheap. Most, if not all, of the growth in smart manufacturing will take place among multibillion dollar companies. Is this money well spent? According to a Capgemini Research Institute survey of 1,000 manufacturers with smart factory initiatives underway, only 14% described their deployments as successful (Baggott, 2020). The A.I. issue in manufacturing is factories aren’t closed, linear systems. They’re chaotic system. Infinitesimal changes in any one of the myriad of input variables to a factory are capable of producing disproportionate changes in output. As a result, no matter how much scrap, downtime, sales or on-time delivery data a company collects actual values will eventually diverge exponentially from what existing A.I. algorithms predict. A.I. will not be capable of running a factory until more research is done on controlling dynamic, nonlinear systems.

Acknowledgements

This paper greatly benefited from input provided by the Strategic Projects and Inbound Customer Engagement group at the SAP Company.

References

- Ang, C. (2020, July 13). Retrieved from Charting the Massive Scale of the Digital Cloud: <https://www.visualcapitalist.com/massive-scale-of-cloud/>
- Anylogic. (2020). *Multimethod simulation modeling for business applications*. Retrieved from file:///C:/Users/seidelsonc/Downloads/multimethod-simulation-modeling-for-business-applications.pdf
- Baggott, A. (2020, Jan 31). *Is it cost-effective to upgrade to a Smart Factory?* Retrieved from <https://www.ued.co.uk/blog/is-it-cost-effective-to-upgrade-to-a-smart-factory-158321/>
- Bryner, J. (2018, June). Retrieved from This Supercomputer Can Calculate in 1 Second What Would Take You 6 Billion Years: <https://www.livescience.com/62827-fastest-supercomputer.html>
- Capgemini Research Institute. (2019). *Smart factories @ scale*. Retrieved from <https://www.capgemini.com/wp-content/uploads/2019/11/Report-%E2%80%93-Smart-Factories.pdf>
- Columbus, L. (2019, Nov. 17). *Smart Factories Will Boost Global Economy \$1.5T By 2023*. Retrieved from <https://www.forbes.com/sites/louiscolombus/2019/11/17/smart-factories-will-boost-global-economy-15t-by-2023/?sh=1f5e5c3f4d81>
- Fan, H. J. (2020, March 30). Long-term prediction of chaotic systems with machine learning. *Physical Review Research*, pp. 1-4.
- Gallaher, M. e. (2016). *Economic Analysis of Technology Infrastructure Needs for Advanced Manufacturing*. Research Triangle Park, N.C.: RTI International.
- Gyarmathy, K. (2020, March 26). *Comprehensive Guide to IoT Statistics You Need to Know in 2020*. Retrieved from <https://www.vxchnge.com/blog/iot-statistics>
- Hao, K. (2020, Dec 11). *Tiny four-bit computers are now all you need to train AI*. Retrieved from <https://www.technologyreview.com/2020/12/11/1014102/ai-trains-on-4-bit-computers/>

- Hessmann, T. (2013, Feb 14). *The Dawn of the Smart Factory*. Retrieved from <https://www.industryweek.com/technology-and-iiot/article/21959512/the-dawn-of-the-smart-factory>
- Huilgol, H. (2017, May 29). Retrieved from Lambda Architecture for Real-time data processing: <https://medium.com/@hsamogh/lambda-architecture-for-real-time-data-processing-62294cdf2b3>
- Industry Week. (2018, June 4). *What's Stopping the Smart Factory Revolution?* Retrieved from <https://www.industryweek.com/leadership/article/22025749/whats-stopping-the-smart-factory-revolution>
- Infoclutch . (2021). Retrieved from <https://www.infoclutch.com/installed-base/erp/>
- Jain, S. a. (2014). VIRTUAL FACTORY REVISITED FOR MANUFACTURING DATA ANALYTICS. *Proceedings of the 2014 Winter Simulation Conference* , (pp. 887 - 898). Washington D.C.
- Knight, W. (2020, July 30). *AI Is All the Rage. So Why Aren't More Businesses Using It?* Retrieved from <https://www.wired.com/story/ai-why-not-more-businesses-use/#:~:text=The%20divide%20between%20the%20two,ones%E2%80%9494are%20using%20the%20te,chnology.>
- Link Labs. (2015, Dec 29). *The Story Behind The Evolution From M2M To Iot To IoE*. Retrieved from <https://www.link-labs.com/blog/m2m-to-iiot-to-iiot>
- Mabkhot, M. A.-A. (2018, June). Requirements of the Smart Factory System: A Survey. *Machins*, 2-23.
- Metz, C. (2017, Oct 22). Retrieved from Tech Giants Are Paying Huge Salaries for Scarce A.I. Talent: <https://www.nytimes.com/2017/10/22/technology/artificial-intelligence-experts-salaries.html>
- Michel, O. a. (1996). Higher Order Statistics for Chaotic Signal Analysis. *Control and Dynamic Systems*, 105-154.
- National Research Council. (1995). Modeling and simulation for the virtual factory. In *Information technology for manufacturing: A research agenda* (p. 113). The National Academies Press.
- Noura, M. A. (2019). Interoperability in Internet of Things: Taxonomies and Open Challenges. *Mobile Netw Appl*, 796–809 .
- Reinsel, D. R. (2020). *Worldwide Global StorageSphere Forecast, 2020–2024: Continuing to Store More in the Core*. Needham, MA.: IDC.
- Sahu, V. (2020, Jan 27). Retrieved from Top 6 Cloud Service Providers who'll dominate the Cloud War: <https://www.techaheadcorp.com/blog/top-cloud-service-providers/>
- SAP Insights. (2020). *What Is Artificial Intelligence?* Retrieved from <https://insights.sap.com/what-is-artificial-intelligence/>
- Saxon, C. (2012, Sept 5). Retrieved from Database Tutorials and Videos: <https://www.databasetube.com/database/big-data-lambda-architecture/>
- Shaw, K. (2019, Nov 13). Retrieved from What is edge computing and why it matters: <https://www.networkworld.com/article/3224893/what-is-edge-computing-and-how-it-s-changing-the-network.html>
- Tan, S. (2019, Mar 28). Retrieved from How to Train Your AI: <https://medium.com/revain/how-to-train-your-ai-98113bdac101>
- Techjury. (2021, Jan 4). *How Many IoT Devices Are There in 2020?* . Retrieved from <https://techjury.net/blog/how-many-iiot-devices-are-there/#gref>
- Techtarget. (2018, Sept). *Storage at the edge*. Retrieved from <https://searchstorage.techtarget.com/definition/storage-at-the-edge>
- Telecom Review. (2020, Mar 5). Retrieved from Living on the edge: Edge computing explained: <https://www.telecomreview.com/index.php/articles/reports-and-coverage/3706-living-on-the-edge-edge-computing-explained>
- U.S. Census Bureau. (2017, Nov 15). *Small Manufacturers Play a Big Role in the U.S. Economy*. Retrieved from https://www.census.gov/newsroom/blogs/random-samplings/2017/11/small_manufacturers.html
- UC Berkeley EECS Dept. (2019). Retrieved from Cyber Physical Systems: <https://ptolemy.berkeley.edu/projects/cps/>
- Ulyanov, A. (2016, July 16). Retrieved from The “Lambda” Architecture for near real-time data processing: <https://cloudgeometry.io/blog/the-lambda-architecture/>
- Wolchover, N. (2018, April 18). *Machine Learning's 'Amazing' Ability to Predict Chaos*. Retrieved from <https://www.quantamagazine.org/machine-learning-s-amazing-ability-to-predict-chaos-20180418/>
- Wu, J. (2020, April 19). Retrieved from Edge AI Is The Next Wave of AI: <https://towardsdatascience.com/edge-ai-is-the-next-wave-of-ai-a3e98b77c2d7>
- Zweben, M. (2016). Retrieved from 5 Reasons Machine Learning Applications Need a Better Lambda Architecture: <https://www.kdnuggets.com/2016/05/5-reasons-machine-learning-applications-lambda-architecture.html#comments>