

Quality 4.0: Artificial intelligence for manufacturing innovation

- Yellow belt training
- Carlos Escobar, PhD
- Quality 4.0 Institute



About the author

Professional experience

- Consultant Quality 4.0 Institute, present.
- Faculty aide, Harvard Extension School, present.
- Postdoc, Tec de Monterrey, present.
- Senior researcher, General Motors – Manufacturing Systems Research Lab, Global R&D, 2015-2021.
- Research assistant, New Mexico State University, 2014-2015.

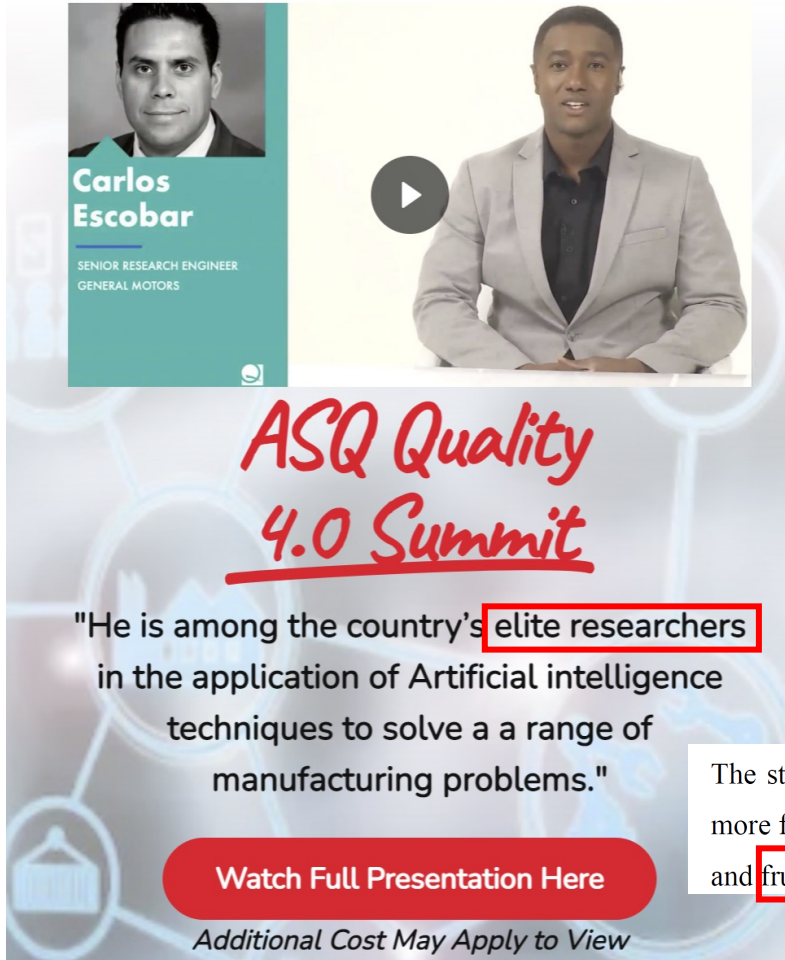
Education

- Master in Management, Harvard Extension School, 2020-2023
- PhD in Engineering Sciences (DCI), Tec de Monterrey, 2019.
- Master in Industrial Engineering, New Mexico State University, 2016.
- Master in Quality Systems, Tecnológico de Monterrey, 2005.
- Ingeniería Industrial, Tecnológico de Ciudad Juárez, 2001.

Scientific production

- 32 scientific publications in top journals.
- Book, Machine Learning in Manufacturing: Quality 4.0 and the Zero Defects Vision, Elsevier, 2023.

Citations and recognition



Carlos Escobar
SENIOR RESEARCH ENGINEER
GENERAL MOTORS

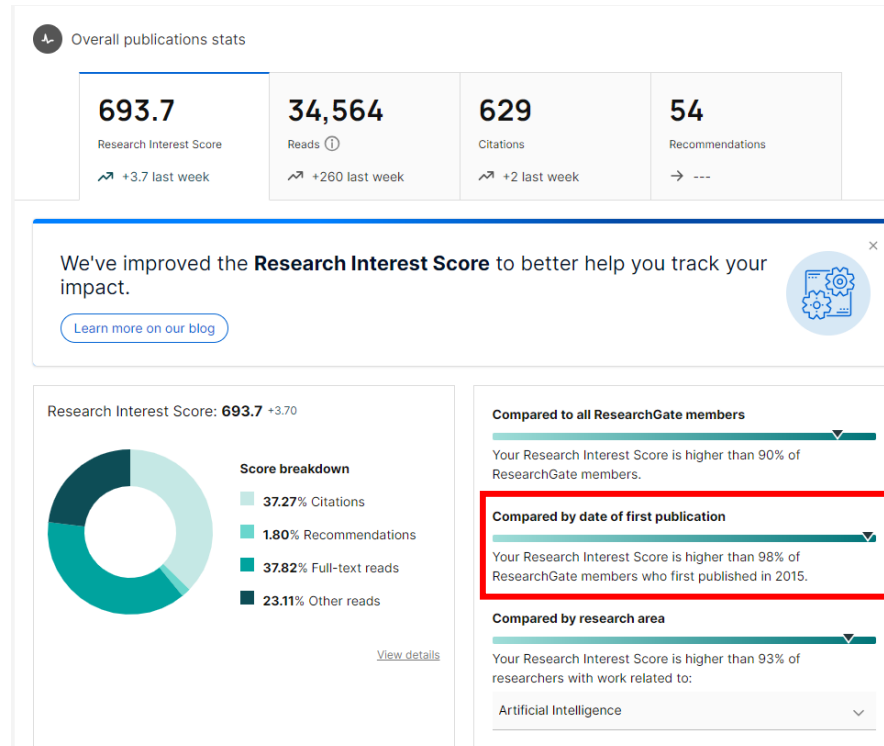
ASQ Quality 4.0 Summit

"He is among the country's elite researchers in the application of Artificial intelligence techniques to solve a range of manufacturing problems."

[Watch Full Presentation Here](#)

Additional Cost May Apply to View

<https://asq.org/conferences/quality-4-0/2021>



<https://www.researchgate.net/profile/Carlos-Escobar-7>

The study revealed Antony, J.; Sony M.; Escobar, C.A.; Morales-Menedez, R. and Nenadáí, J. as more fruitful authors and Antony, J. and Sony, M. as the two most cited ones. The top 19 most cited and fruitful authors are presented in the Table 1.

Quality 4.0: a bibliometric study and research agenda August 2022
Conference: 5th International Conference on Quality Engineering and ManagementAt: University of Minho, Braga, Portugal



INVESTIGACIONES DE ALTO IMPACTO

CARLOS ESCOBAR,
líder de
Process
Monitoring
for Quality

PROCESS MONITORING FOR QUALITY
El valor de la calidad

SECTOR: MANUFACTURA

El fotógrafo Robert Frank lo tenía claro: "lo importante es ver aquello que resulta invisible para los demás". Esa es la esencia de la investigación que, en 2016, inició Carlos Escobar para detectar los defectos que las empresas de manufactura no identifican por seguir las técnicas analíticas tradicionales en sus procesos de producción.

"Todo empezó cuando estudié mi maestría en Calidad y Productividad en el Tec de Monterrey", recuerda el investigador senior de los laboratorios globales de I+D de General Motors. "Estuve expuesto a los contextos de la aplicación estadística de la calidad, que no era tan común. No estaba explotada la inteligencia artificial".

Se especializó y, como parte de sus estudios doctorales, abordó los retos teóricos y prácticos para desarrollar una filosofía de calidad basada en big data e inteligencia artificial. Esto permitió que las compañías detecten defectos que podrían pasar desapercibidos.

El propósito de Escobar es generar valor a través del control de calidad. Sabe que desarrollar procesos que generen cero defectos es tan relevante como asegurarse que un marcapasos funcione.

https://issuu.com/tecreview/docs/tec_review-30





Citations and recognition

“Dr. Escobar Diaz is constantly innovating and solving tough problems, as proven by his numerous technical successes in manufacturing research. He performs world class caliber research...”

Tony Argote
Director, Autonomous Technology at General Motors

Filminas están en ingles.

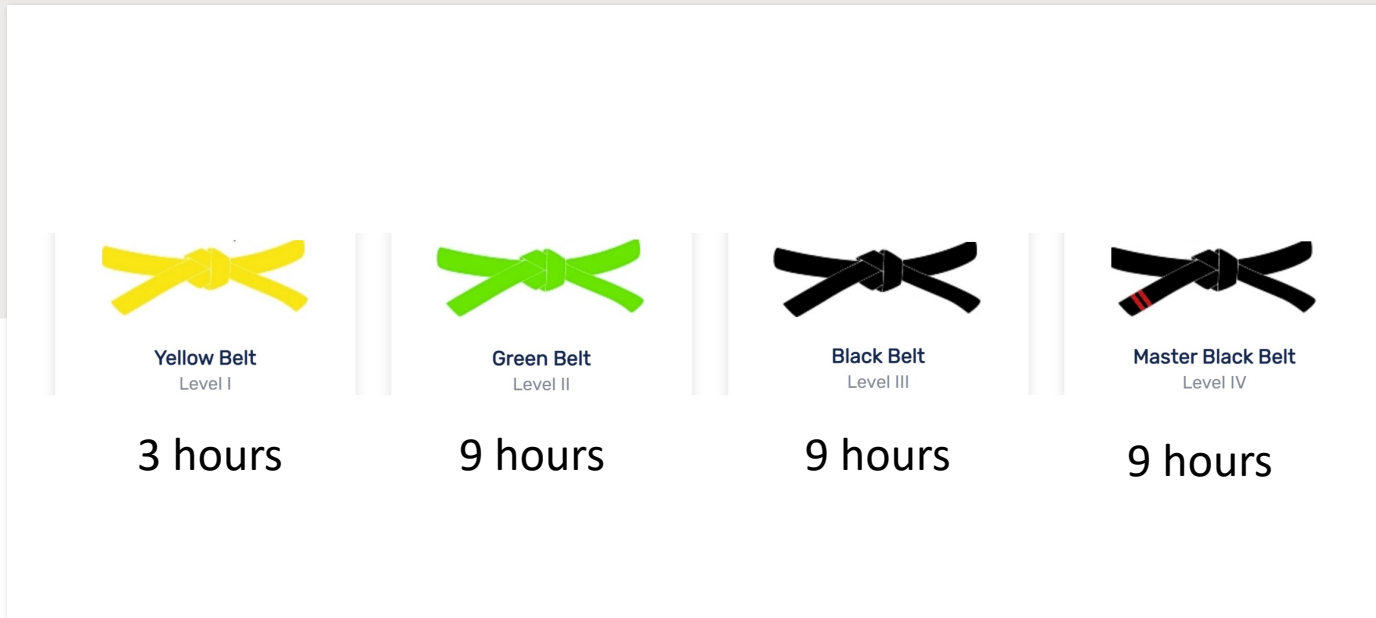
Presentación, preguntas, y discusiones en español.

Algunas frases y/o términos se describiran en ingles.

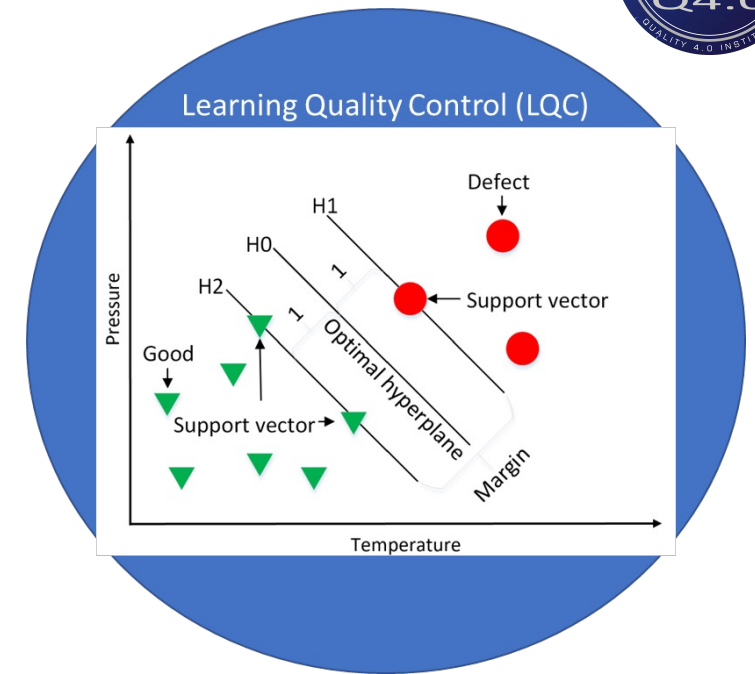
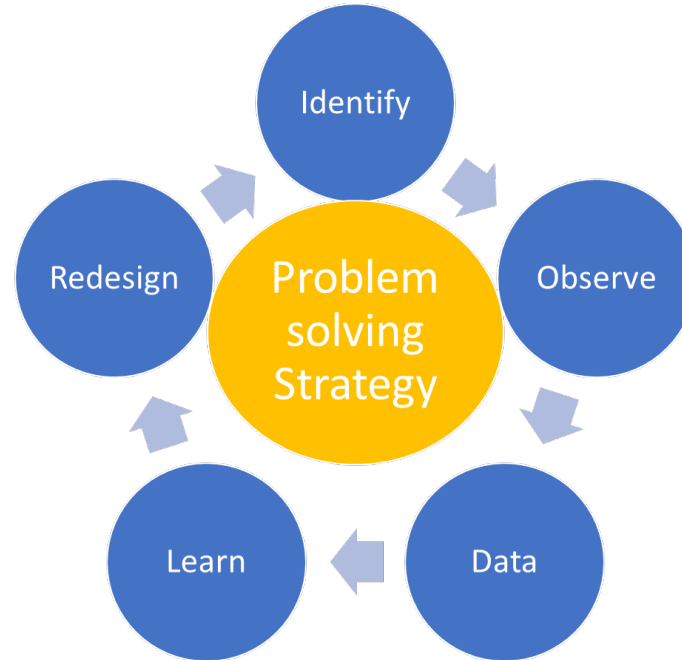
Preguntar en cualquier momento.

Notes:

About the program



- Mastery level defined by karate convention
- 30 hours of training
- Curricula based on the author's book
 - Theory, empirical evidence, and the author's experience
- Six areas of knowledge (statistics, programming, optimization, learning, manufacturing, quality)
- Problem solving strategy
- Learning quality control (LQC) system



About the program

Approach

These series of courses will help:

About the program

Managers to develop a vision for Quality 4.0

Engineers to develop and implement intelligent quality systems

Math and coding aspects at levels that are easily understandable for most engineers

Focus more on the practical aspects of selecting and solving engineering intractable problems through AI

About the program

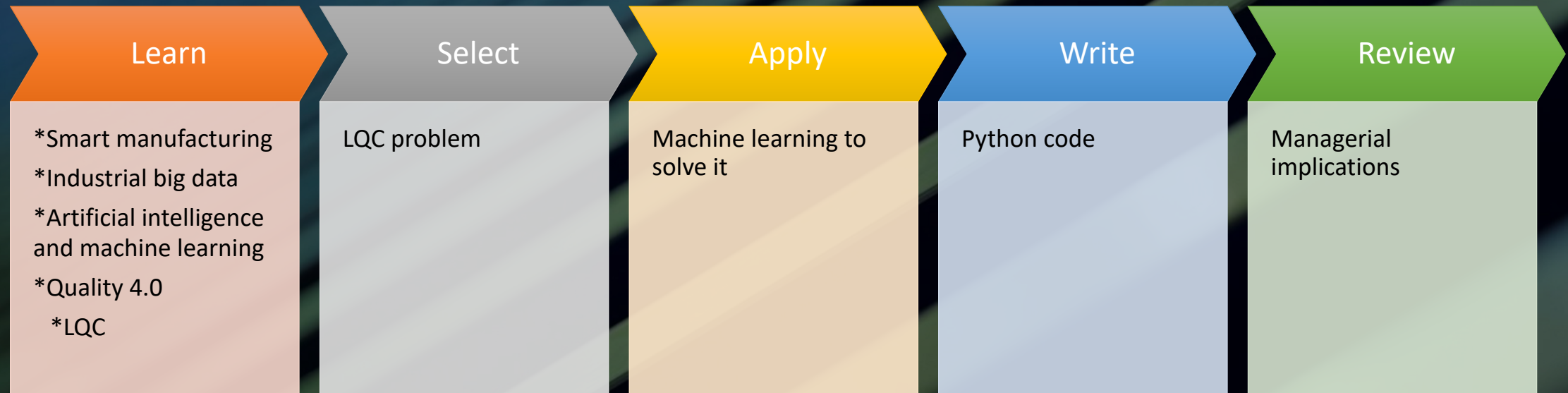


Quality 4.0 Institute

www.Quality4.com
info@Quality4.com

 YELLOW BELT	 GREEN BELT	 BLACK BELT	 MASTER BLACK BELT
Gain the Knowledge to Drive Innovation in Your Organization			
• Smart Manufacturing ◦ Technologies ◦ Building blocks ◦ Characteristics ◦ Research areas • Manufacturing Big Data ◦ 10V's ◦ Data sets for binary classification of quality • Evolution of Modern Quality Control ◦ Problem solving strategies and paradigms • Breakdown of Traditional Quality Control Methods ◦ Trends ◦ Case Study ◦ Limitations • The Rise of Quality 4.0 ◦ Definitions ◦ Objectives ◦ Areas of Knowledge • The Technologies ◦ Artificial Intelligence ◦ Cloud Storage & Computing ◦ Industrial Internet of Things & Cyber Physical Systems • Introduction to Project Selection	• Binary Classification of Quality ◦ Machine learning theory ◦ Pattern analyses • Learning Quality Systems ◦ Definition ◦ Applications • Evolved Problem Solving Strategy ◦ Identify ◦ Observe ◦ Data ◦ Learn ◦ Redesign • Python Installation • Public Data Sets Access • Linear Classification ◦ Support vector machine • Project Selection Advise	• Data Preprocessing Techniques ◦ Visualization ◦ Missing Values ◦ Imputation Methods ◦ Feature Scaling ◦ Outlier Analysis ◦ One Hot Encoding • Feature Selection ◦ Filter methods ◦ Wrapper ◦ Embedded • Non-linear Classification ◦ Xgboost ◦ Random forest • Model Selection • Project Review	• Deep Learning ◦ Feed Forward Neural Network ◦ Convolutional Neural Network • Meta-Learning ◦ Multiple Classifier Systems • Managerial Implications ◦ Team Building ◦ Project Selection Theory ◦ Outsourcing Evaluation ◦ Pilot Runs ◦ Getting the Most From AI • Prediction Optimization Techniques • Project Preliminary Results Analysis • Algorithm Development (optional) • Publication Support (optional)

Full program learning objectives



Requirements for certification

- Theory course
- One project per program
 - Yellow – define two to three candidate projects
 - Green – project selection, data generation, linear classifier
 - Black – non-linear classifier, feature relevance analysis
 - Master black – deep learning, process redesign, practical and managerial implications
- Project report
 - PPT for yellow, green, and black belts
 - Paper for mater black belt



Yellow belt
training



Did you know?

- Industrialization of AI is today a megatrend that dominates the business landscape
 - 92% of the surveyed leaders are increasing their investments in big data and AI
- AI technologies can automate 40% of worldwide workforce tasks. The return-on-investment argument has been substantially demonstrated
- Lack of powerful people is one of the biggest challenges facing these technologies in business
 - Leaders are universally citing difficulties in developing a vision for AI
- Machine learning engineer is the best job in the U.S.

Sources:

- LLC NewVantage Partners. Big Data and AI Executive Survey 2019: Data and Innovation How Big Data and AI are Accelerating Business Transformation, 2019.
- Laurence Goasduff. 2 megatrends dominate the gartner hype cycle for artificial intelligence, 2020. Gartner, September 2020.
- A.I. Expert Says Automation Could Replace 40% of Jobs in 15 Years, Bydon Reisinger, Fortune, January 10, 2019.
- The AI talent shortage isn't over yet, David Jarvis, Deloitte, 30 September 2020.
- Machine Learning Engineer Is The Best Job In The U.S., Indeed, Louis Columbus, Mar 17, 2019.

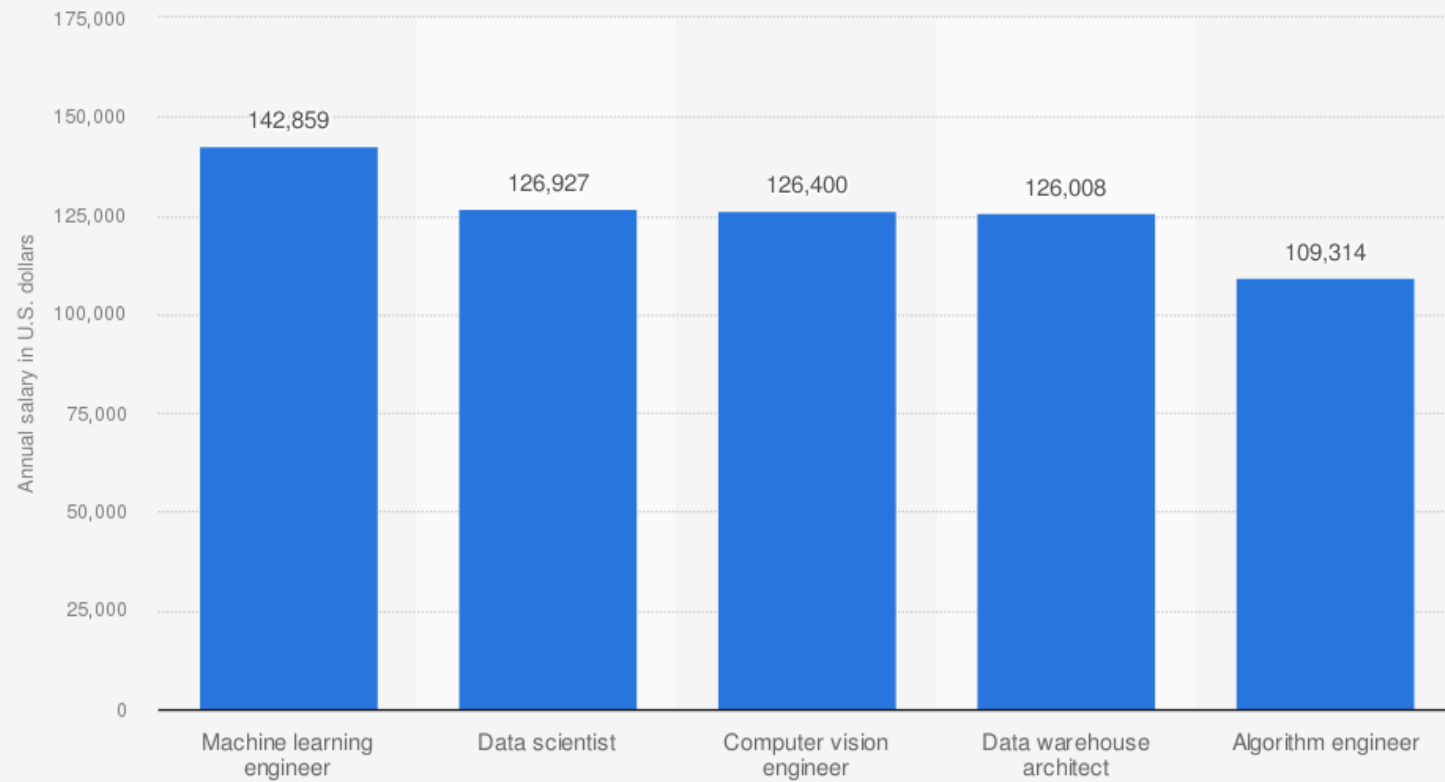
Did you know?

- Better-educated, better-paid workers (along with manufacturing and production workers) will be the most affected by the new AI technologies
- Quality control is the most cited priority for the application of this technology
- Around 50% of quality inspections are mostly manual—less than 10% automated
- AI improves product quality by reliably detecting defects and speed by automating more steps in the process that are traditionally done visually
- Salary of certified professionals is directly related to their color belt
- Around 60-73% of their data remains dark; unanalyzed, unorganized, and unused

Sources:

- What jobs are affected by AI? Better-paid, better-educated workers face the most exposure. Brookings, Mark Muro, Jacob Whiton, and Robert Maxim November 20, 2019.
- The artificial intelligence effect on industrial products. IBM Institute for Business Value. José R Favilla, et. Al., 2018.
- Automation Opportunities Abound for Quality Inspections. Automation World. Michael Belfiore. Jun 16, 2016.
- How AI Software Enables More Accurate Quality Inspection. Quality Magazine. Bruno Menard. May 5, 2021
- Up to 73 percent of company data goes unused for analytics. Inc., Jeff Barret. April 2018.

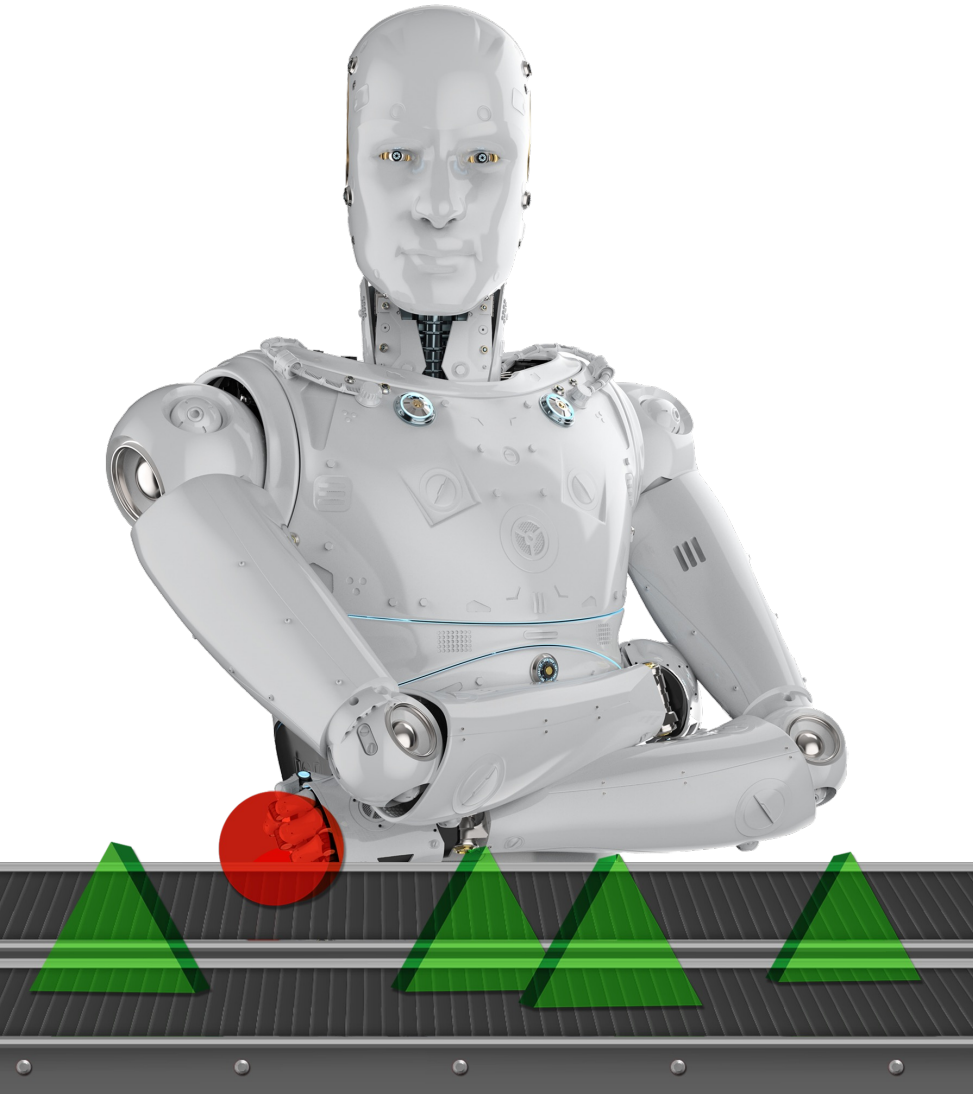
**Artificial intelligence (AI) job titles with the highest salaries in United States as of 2019
(in U.S. dollars)**



Source
Indeed
© Statista 2022

Additional Information:
United States; 2019

Statista User-ID: 5217289

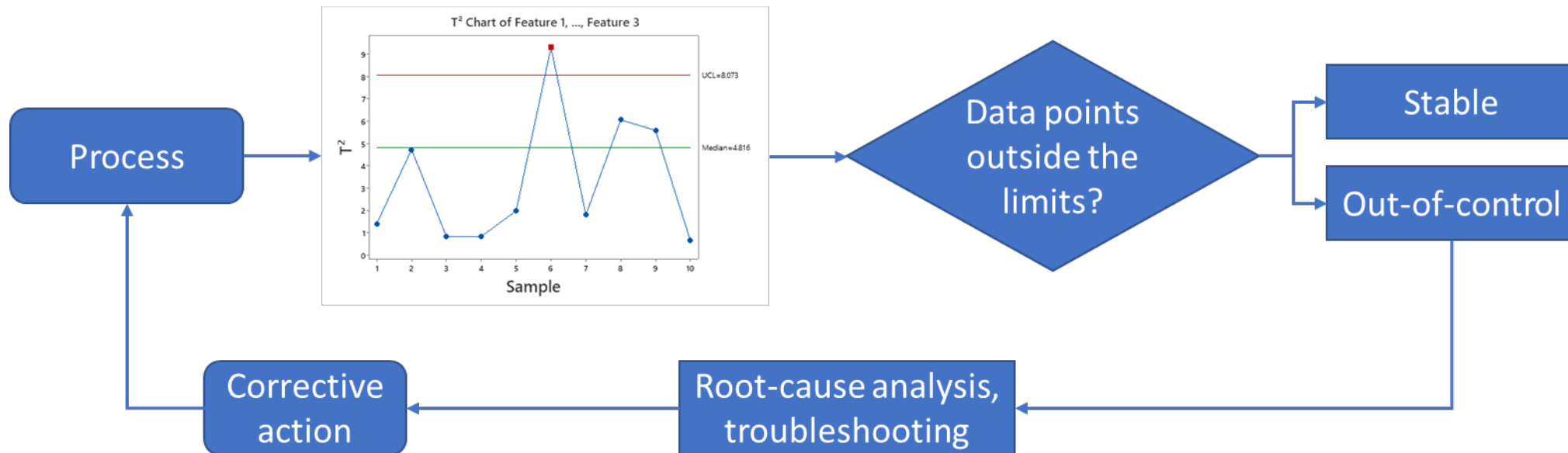


Quick intro

Learning quality control (LQC) the evolution of statistical process control (SPC)

SPC control charts analyze process performance by plotting data points, control limits, and a center line.

Common control charts: range (R), standard deviation (S), moving range (MR), mean (X-bar), and individuals (I) charts



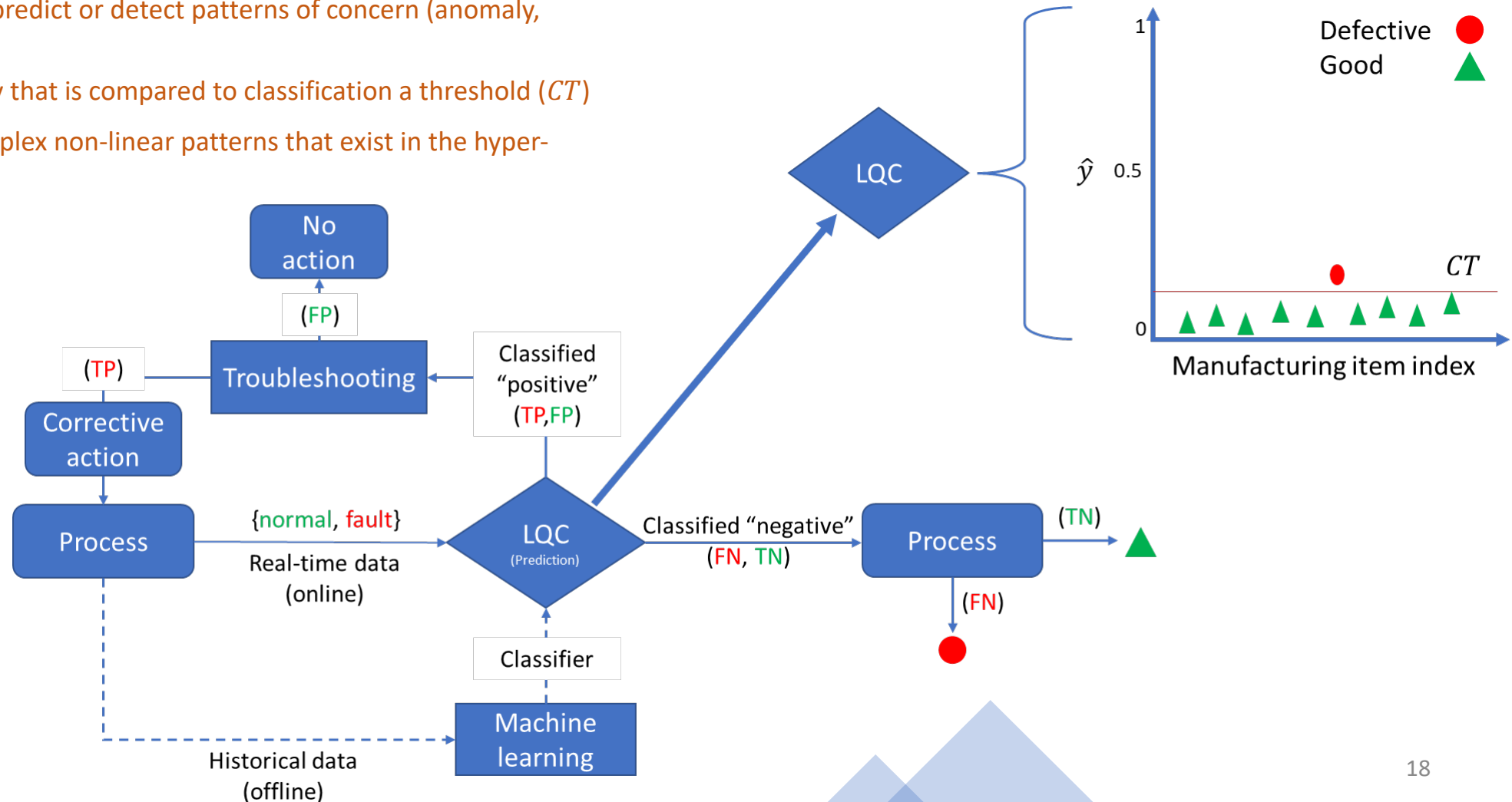
Learning quality control (LQC) the evolution of statistical process control (SPC)

LQC, machine learning algorithms (MLAs) are applied to historical data to develop a classifier with the capacity to predict or detect patterns of concern (anomaly, deviation, defect)

Classifier generates a probability that is compared to classification a threshold (CT)

MLAs are better at learning complex non-linear patterns that exist in the hyper-dimensional space

LQC boost SPC



Chapter 1: Introduction

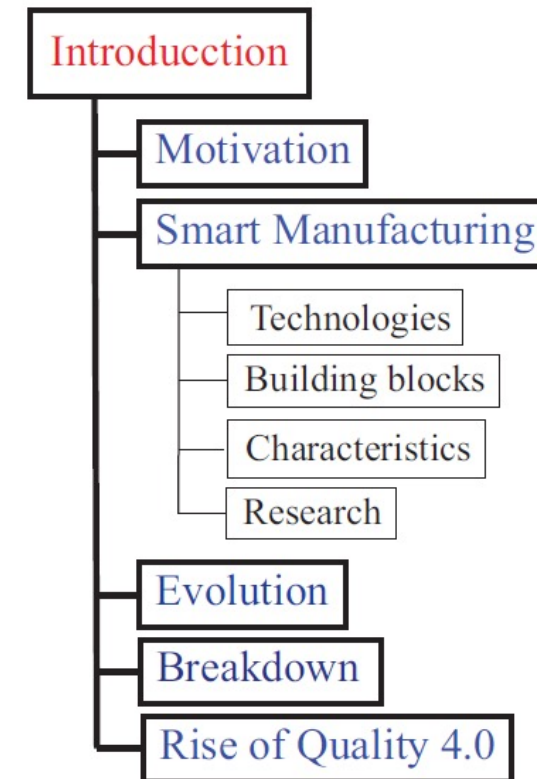
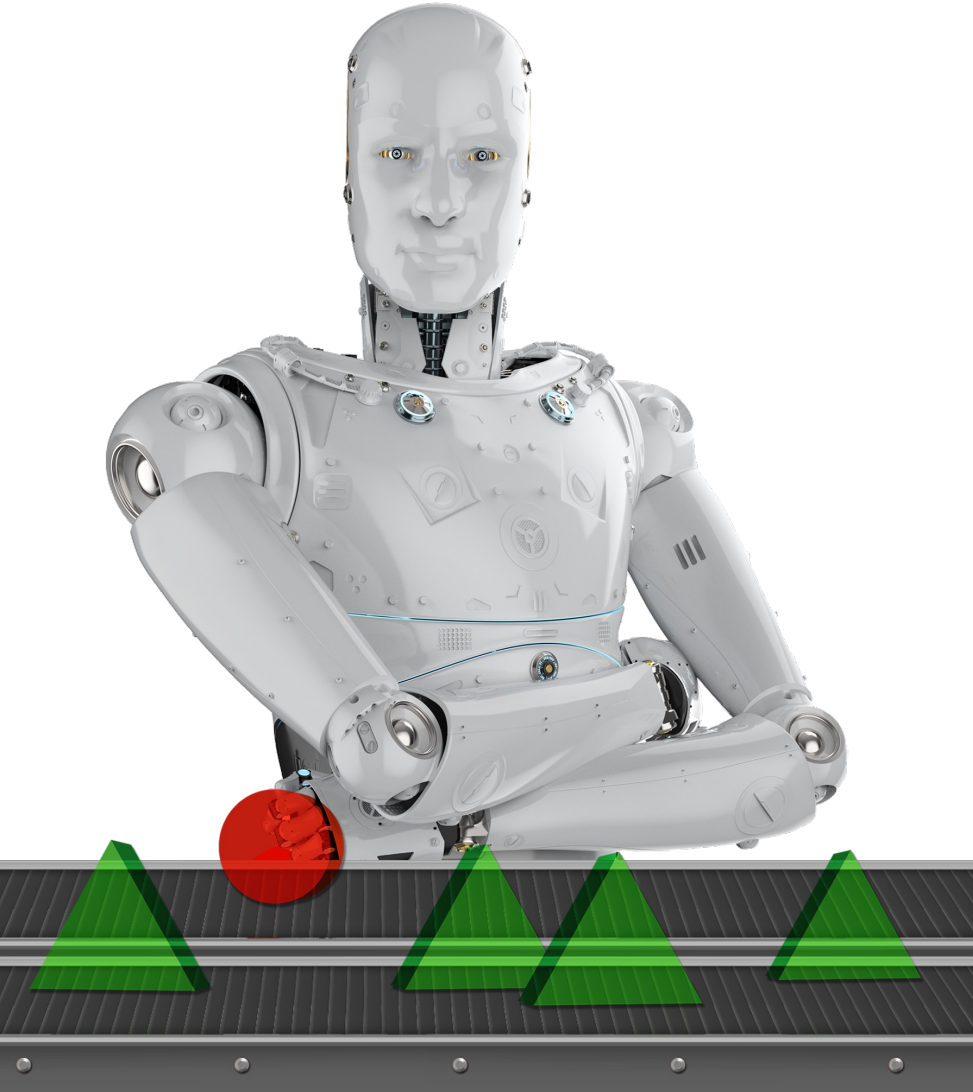


FIGURE 1.1 Contents of the Chapter 1

Motivation, importance of manufacturing

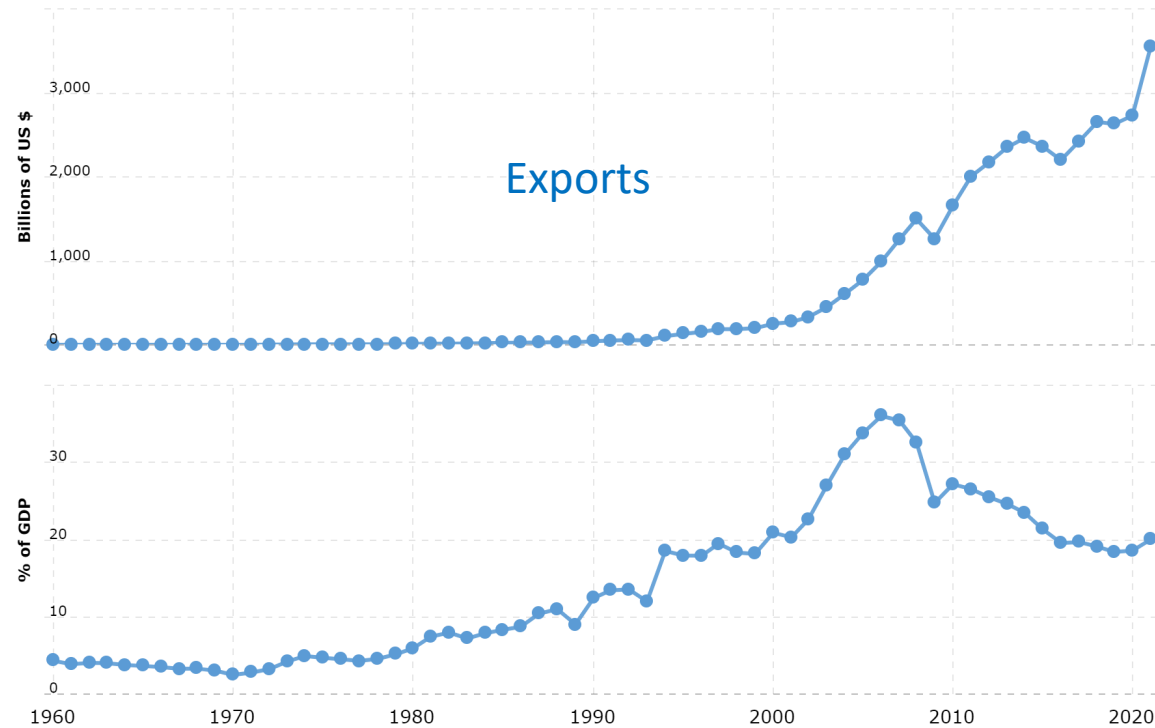


Importance of manufacturing

- Manufacturing is the driving economic force of the most advanced countries
- Global market for manufacturing is forecasted to grow from US \$ 649.8 billion in 2020 to US \$ 732.2 billion in 2027
- Manufacturing increases the living standards by increasing the purchasing power of the people living in industrialized societies
- Advanced manufacturing companies are leaders in innovation, productivity, exports, and research and development
- Most technological advances have originated in these companies
- For every dollar of domestic manufacturing value-added, another \$3.60 of economic activity is generated elsewhere across the economy
- For every manufacturing job, there are 3.4 jobs created in non-manufacturing industries

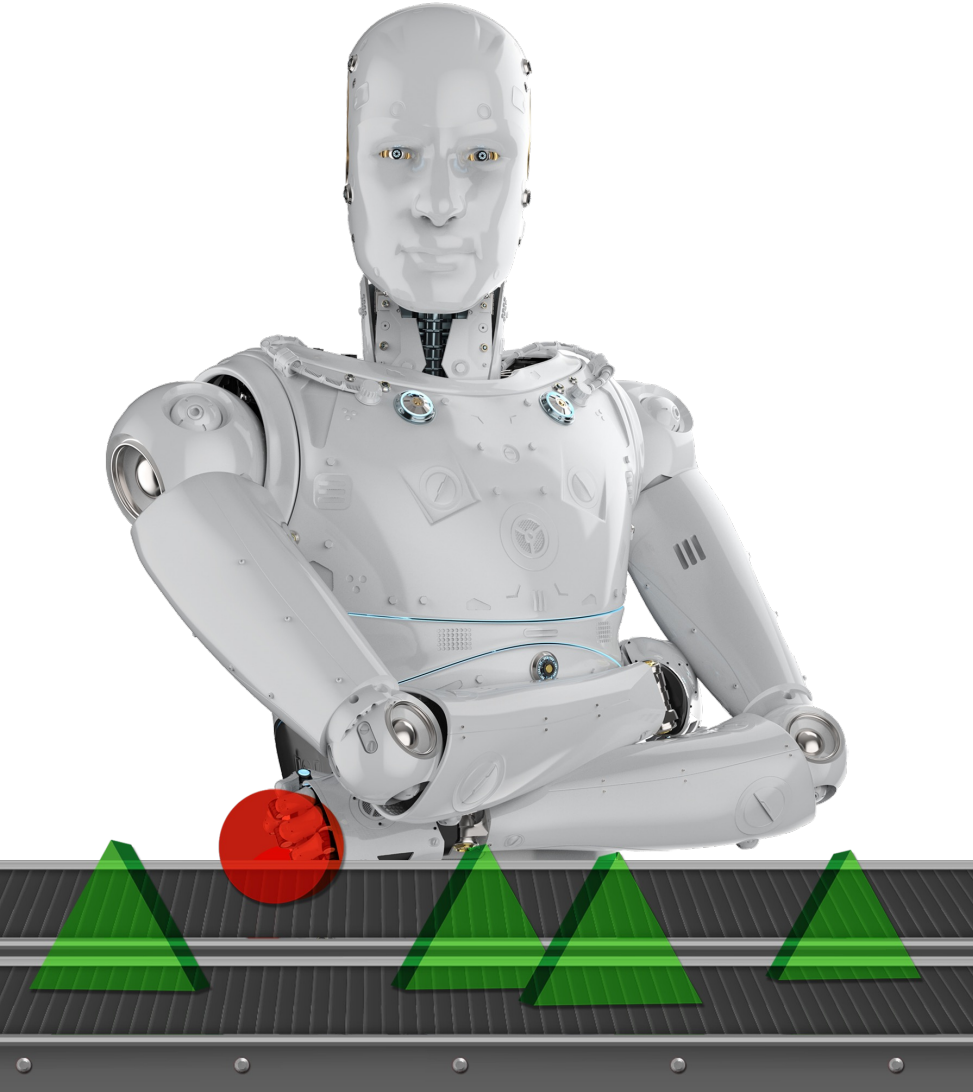
Importance of manufacturing

- China is the world's largest manufacturing economy
- Manufacturing helped China to rise as a global economic superpower



Source: <https://www.macrotrends.net/countries/CHN/china/exports>

Motivation, importance of quality



Importance of quality

- High quality products is a top priority for most manufacturing companies.
- High quality improves customer retention, builds brand trust, and boosts return-on-investments, enabling business growth

Definitions

- “Product or service that is free of deficiencies”, ASQ
- “Fitness for use”, *Joseph Juran*
- “Eight dimensions: (1) performance, (2) features, (3) reliability, (4) conformance, (5) durability, (6), serviceability, (7) aesthetics, and (8) perceived quality”, *David Garvin*
- “Conformance to requirements”, *Philip Crosby* ← Most relevant definition

Importance of quality

Four decades ago, Crosby introduced the concept of zero defects in manufacturing (ZDM)

- Creation of unrealistic standards was criticized by *E. Deming and J. Juran*
- “Eliminate slogans, exhortations, and targets for the work force asking for zero defects and new levels of productivity”, *E. Deming*
- *Crosby* replied, “It is merely setting performance standards that no one can misunderstand”

Importance of quality

Quality systems

- Quality control (QC), process of applying statistical ~~–, machine learning, –~~ and analytical techniques to determine if a manufactured item conforms to the design specifications
- Quality Assurance (QA), planned and systematic activities implemented within the quality system that can be demonstrated to provide confidence that a product or service will fulfill the quality requirements
- Quality Improvement (QI), projects aimed at creating breakthrough levels of performance by eliminating defects

Importance of quality

- QC, QA, & QI created the high conformance production era
- Modern manufacturing systems have ensured that defects are rarely generated
- Most manufacturing companies operate at four sigma (4σ) level (6210 DPMOs)

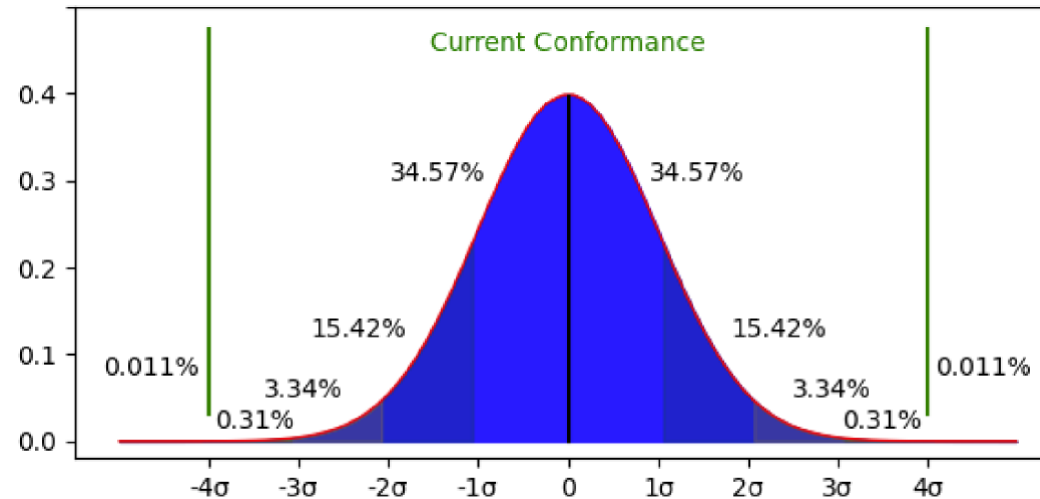


FIGURE 1.8 Four sigma conformance rate.

However,...

Importance of quality

- Customers expect perfect quality
- Intense global competition has led to low profit margins
- Warranties can make the difference between profit and loss
- Customers now use the Internet and social media tools (e.g., Google product review) to share their product experiences
- Reviews can go viral in a couple of days or even hours
- Organizations have little flexibility to recover from their mistakes
- Single negative customer experience can immediately affect a company's reputation



inkReliability®

Cause Mapping® Training Services Tools & Resour

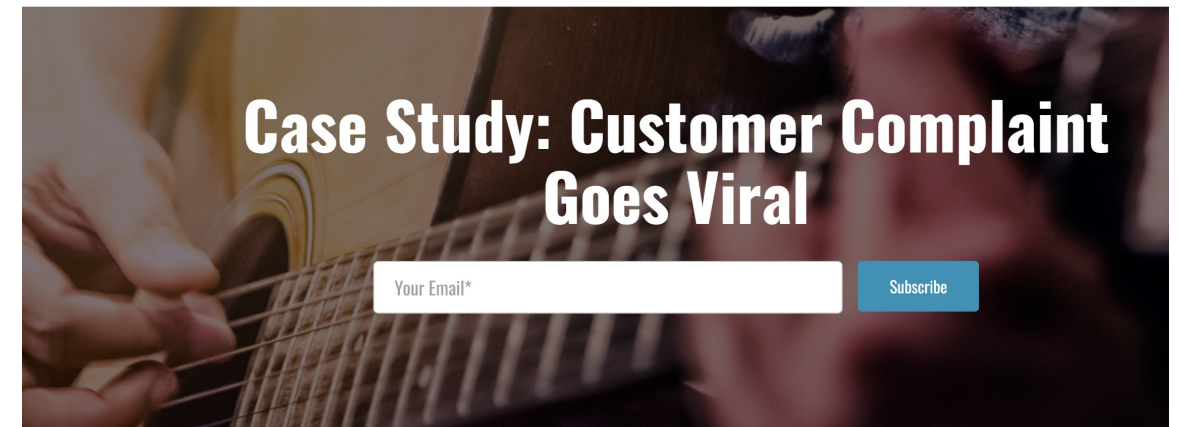
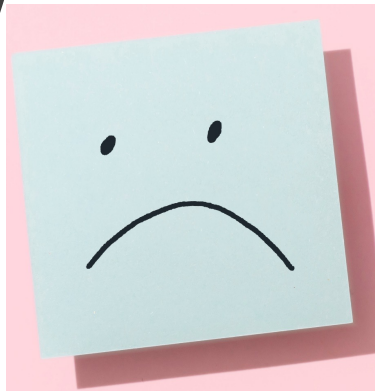


Image source: <https://blog.thinkreliability.com/case-study-customer-complaint-goes-viral>

Importance of quality



Kim Smiley

Nov 4, 2019

Category: Cause Mapping

In 2009, Dave Carroll's Taylor acoustic guitar was damaged after it was checked on a flight from Halifax, Nova Scotia to Nebraska. Carroll and his band were traveling to Nebraska for a one-week tour and checked their guitars because the size and shape of the instruments made them difficult to carry on. Carroll claimed he saw baggage handlers throw guitar cases while loading luggage during a layover at Chicago O'Hare International Airport. Once he landed in Nebraska, Carroll discovered his \$3,500 guitar had been damaged significantly—requiring a \$1,400 repair.

Compensation for Damage Denied

Carroll contacted the airline's customer service department for compensation, but was unable to get reimbursed for repairing his guitar. The airline, United Airlines, didn't deny the damage occurred but said he had failed to process his claim within a 24-hour window. Carroll continued to call and email various customer service personnel, and everyone said there was nothing they could do to help him. After nine months, Carroll was given a final "no" from the airline.

Music Video Goes Viral

Carroll's frustration must have been inspirational because he wrote a song called "United Breaks Guitars Song 1," the first in what would become a trilogy of songs about the incident posted on YouTube. The song is catchy, and the video is funny without being too bitter. But under the humor, the memorable lyrics are the stuff public relations nightmares:

*"You broke it, you should fix it
You're liable, just admit it.
I should've flown with someone else
Or gone by car
'Cause United breaks guitars."*

More than 3 million people viewed the video within 10 days of its release. To date, the original video has more than 19 million views. The other two songs never reached the popularity of the first one, but the second video has over 2 million views and the third video has over a million.

Airline Response

After the video reached 150,000 views, United Airlines contacted Carroll and offered to compensate him for the damages. He told them to donate the money to charity, which they did. The airline also called and apologized for the incident and asked permission to use the video for internal training, so they could work on improving their customer relationships.

But the damage to United's image was done. The media quickly picked up the story of the viral revenge

Importance of quality

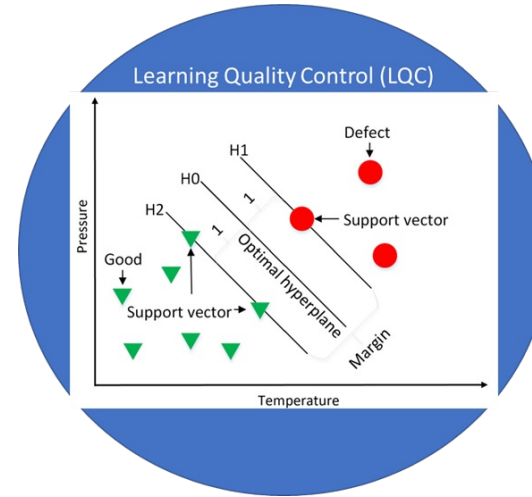
Zero defect manufacturing – from a customer perspective

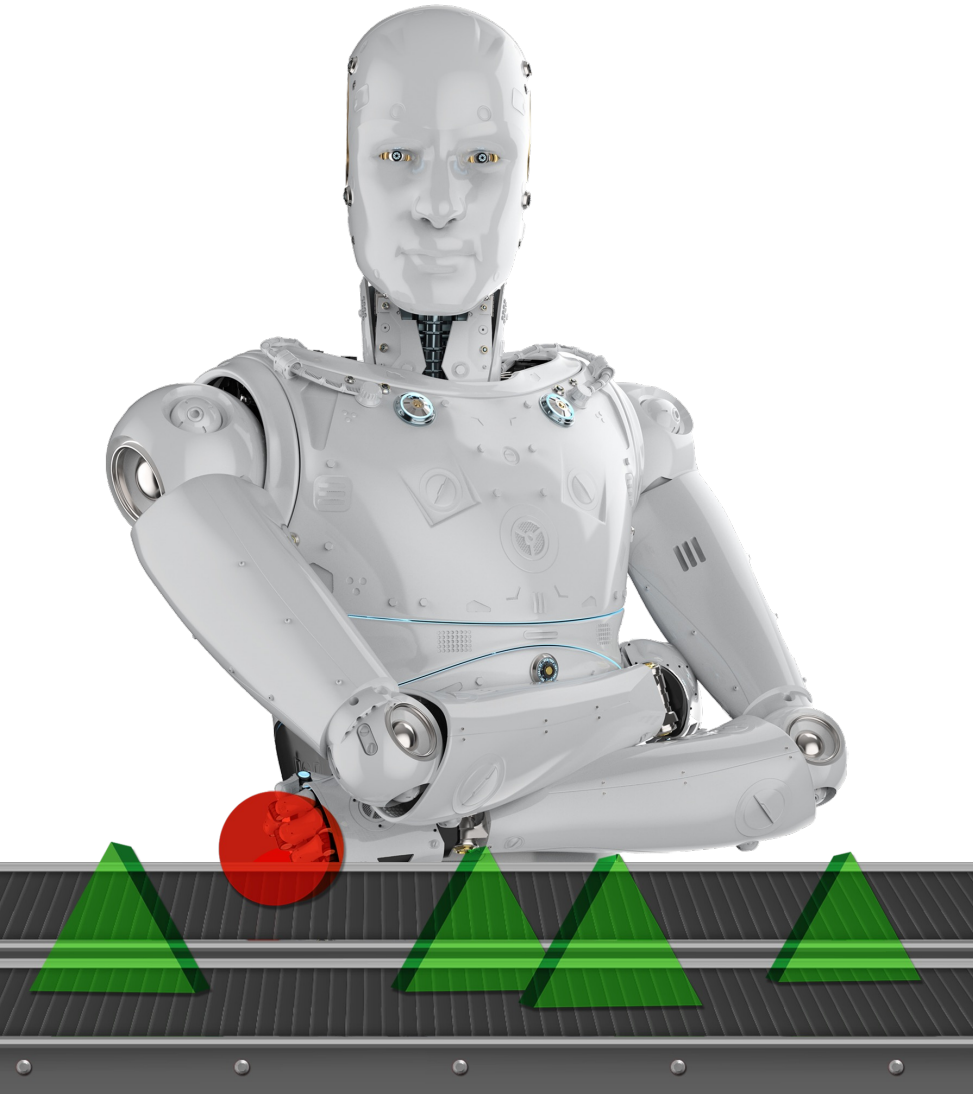
1. Perfect process execution
2. Perfect defect detection



Importance of quality

- Introduction of Artificial Intelligence (AI) into the manufacturing process vastly improves upon the zero-defects vision promoted by Crosby more than four decades ago
- LQC improves upon the zero-defects vision from a technological perspective





Smart manufacturing (SM)

Smart manufacturing

Definition

“leveraging digital transformation, through the use of new Industry 4.0 technologies, within Manufacturing to drive performance, increase quality, reduce cost and scrap, improve reliability and agility, and save energy”, *Smart Manufacturing Innovation Institute (CESMII)*

“a vision fully-integrated, collaborative manufacturing systems that respond in real-time to meet changing demands and conditions in the factory, in the supply network, and in customer needs”, *National Institute of Standards and Technology (NIST)*

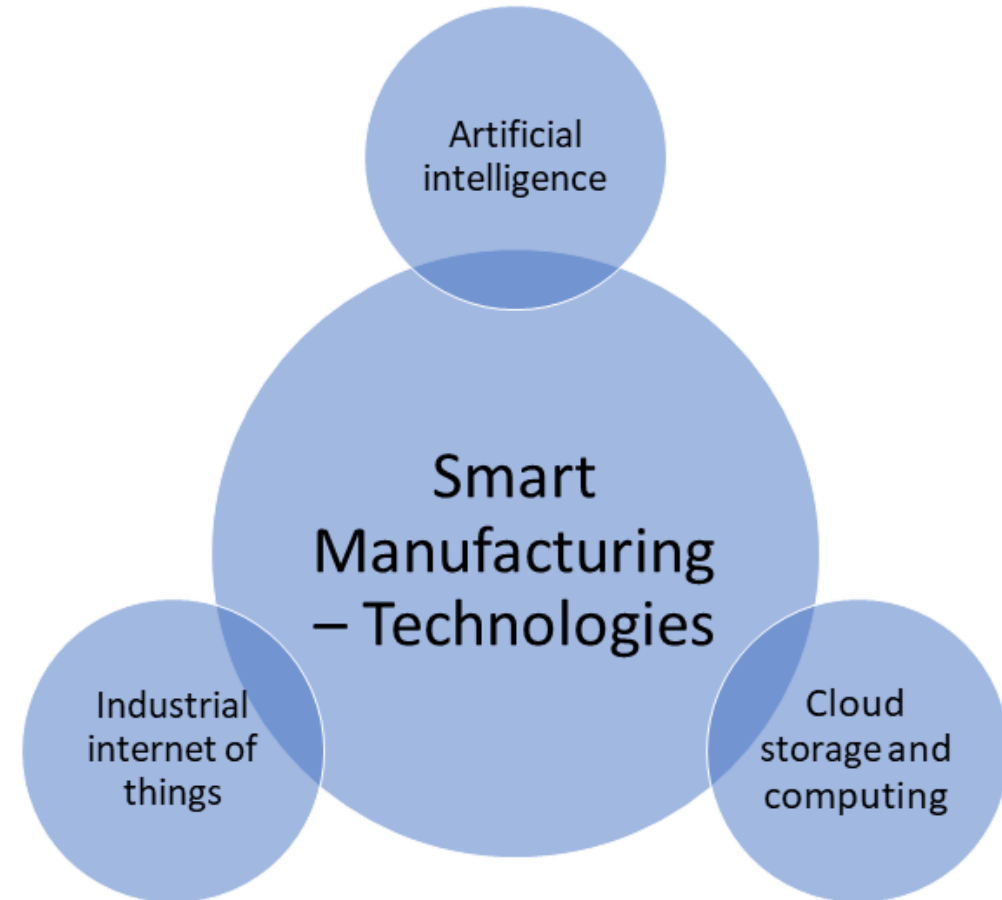
Smart manufacturing

- Term SM was initially introduced at the National Science Foundation workshop on cyberinfrastructure in 2006
- Helps manufacturers increase their efficiency, stay ahead of competition, and explore new business models and practices
- SM has gained interest worldwide and is part of the comprehensive economic development strategy of several countries
 - Industry 4.0 initiative, *Germany*
 - Made in China 2025, *China*
 - Standards Landscape for Smart Manufacturing, *United States*
 - Smart Industry Readiness Index (SIRI), *Singapore*

Smart manufacturing

Technologies

- AI, IIoT, CSC
- New disruptive technologies of the Fourth Industrial Revolution are moving forward the frontiers of manufacturing sciences
- Previous industrial revolutions, most of the value was created by upgrading manufacturing assets
 - Value and innovations in this revolution are not necessarily linked to major machinery upgrades
 - Innovation comes from the cognitive computing capabilities that drive disruptive technologies

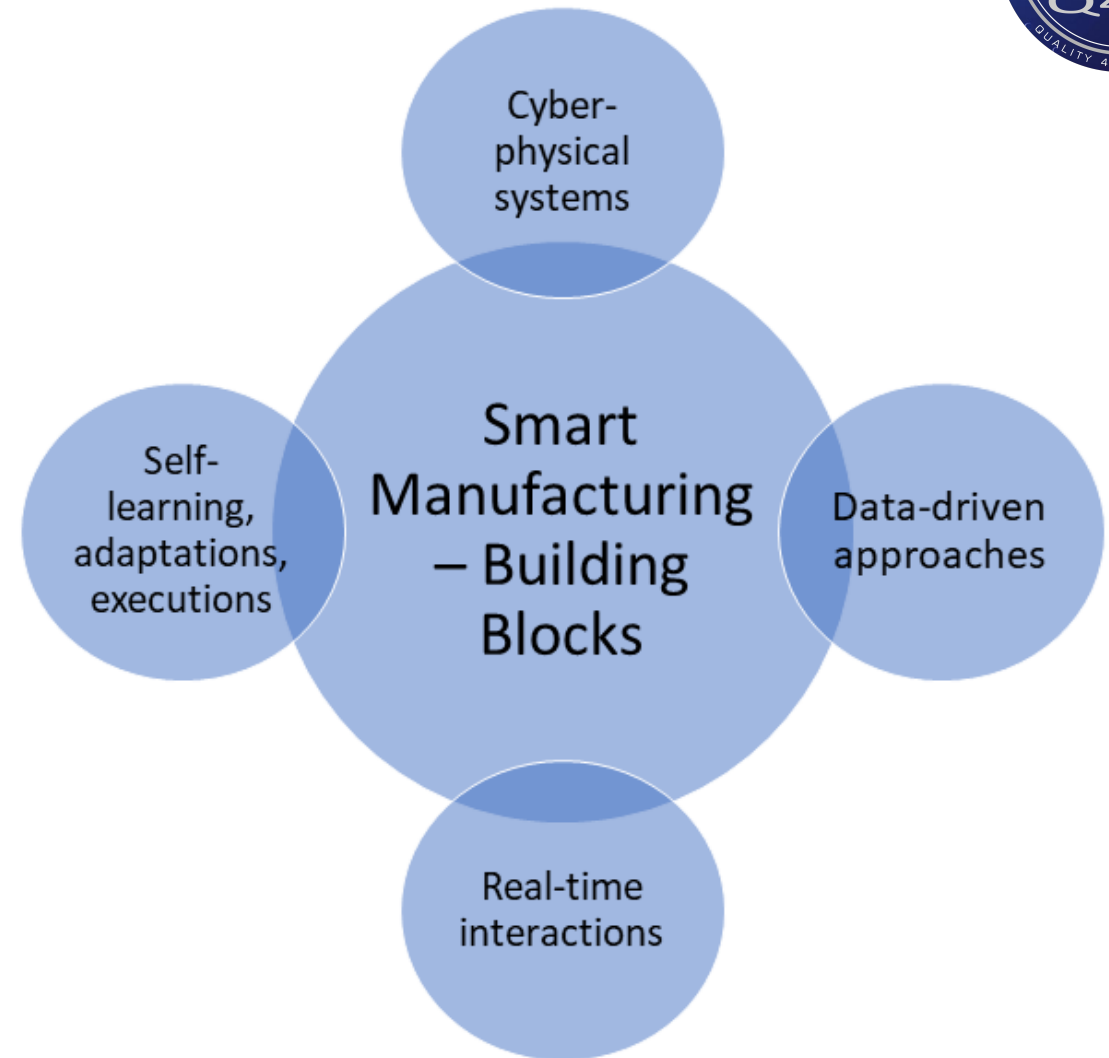


Their combination enables a smart and connected manufacturing environment

Smart manufacturing

Building Blocks

- Cyberphysical Systems (CPS)
- Data-driven approaches
- Real-time iterations
- Self learning adaptations and executions

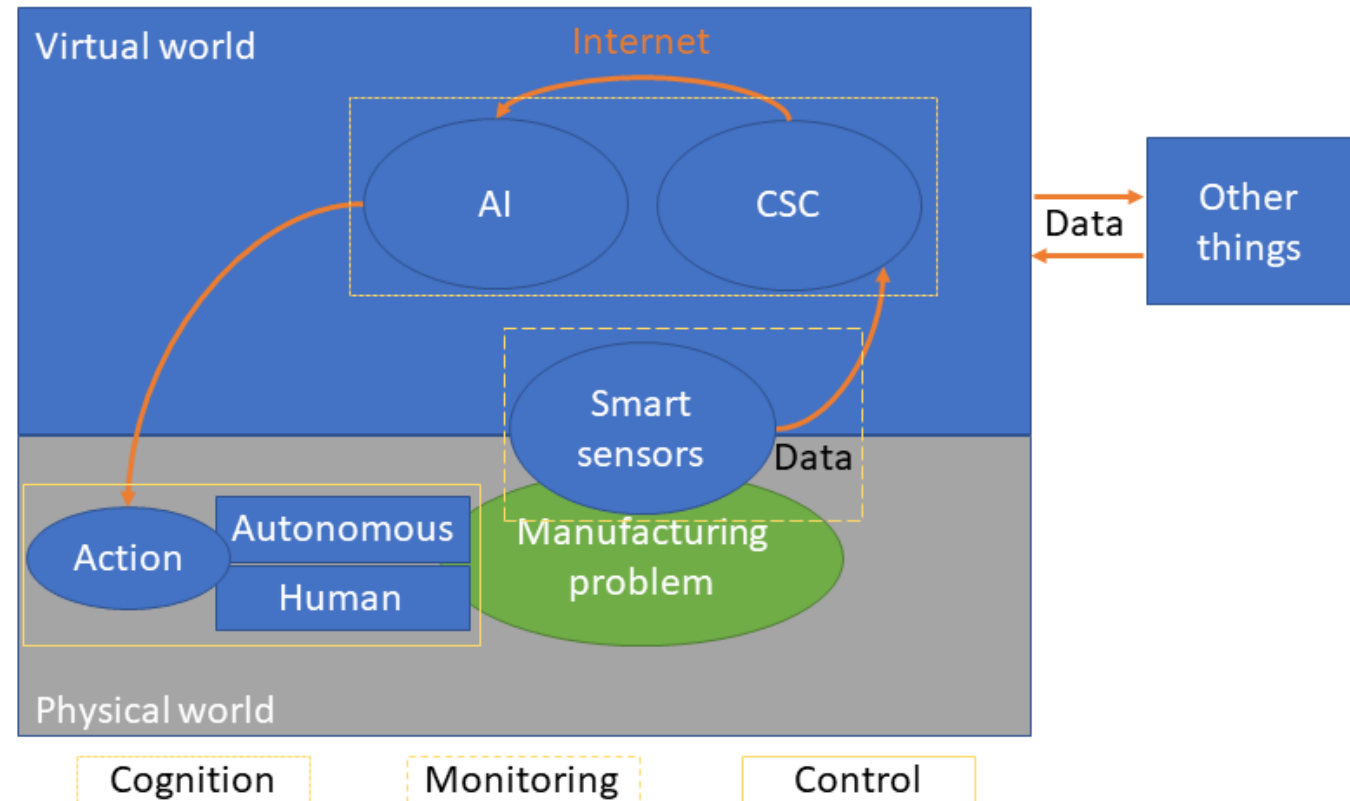


Smart manufacturing

Cyberphysical Systems (CPS)

- Integrates the technologies to fuse the virtual world with the physical world to develop the cognition of SM systems
- “CPS are integrations of computation, networking, and physical processes. Embedded computers and networks to monitor and control the physical processes, with feedback loops where physical processes affect computations and vice versa”

Source: S. Shyam Sunder and Edward A. Lee. Cyber-physical systems, 2021



Activity 1

Access the following links and derive 2-3 remarks, 10mins

- <https://www.linkedin.com/pulse/hidden-risk-ai-big-data-vegard-flovik-phd>
- <https://towardsdatascience.com/observational-vs-experimental-study-543425a3b3c8>
- <https://www.forbes.com/sites/kalevleetaru/2019/01/15/a-reminder-that-machine-learning-is-about-correlations-not-causation/?sh=147daf516161>

Discussions, 5mins

Smart manufacturing

Data-driven approaches

- Observational data
 - Collected data based on what is observed
 - Observational studies are less expensive than experimental studies
 - Researchers do not impose any kind of treatment to the subjects or manipulate variables
 - Subjects are not randomly assigned to a group
 - Only relationships or associations may be defined
 - Not generated to find true cause-and-effect relationships
 - Correlation does not mean causation
- Experimental data
 - Randomly assign a treatment to a group
 - Researchers can draw the cause and effect (causal) conclusion
 - Evidence provided by an experimental study is stronger than the observational study
 - Experimental study involves two groups: (1) experiment group and (2) control group.
 - Control group is the group in the study that does not receive treatment

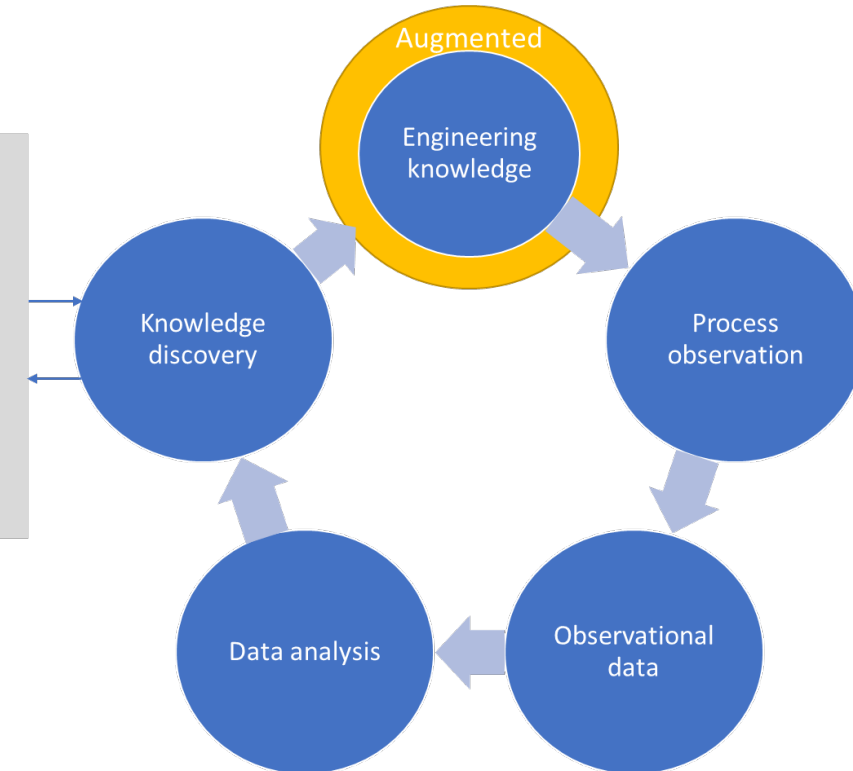
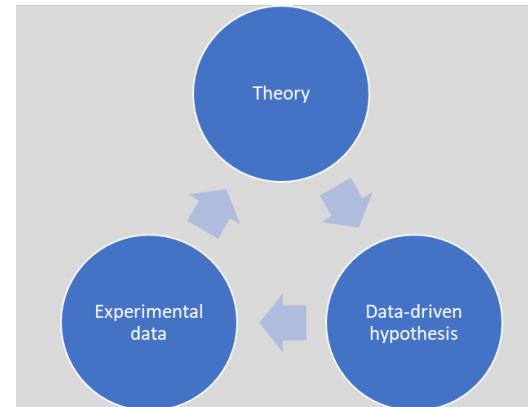


Smart manufacturing

Data-driven approaches

- New generation of manufacturing is changing from engineering-based to a combination of data-driven and engineering-based approaches
- Preconceived engineering notions are used to observe the process using connected devices. They generate high volumes of relevant observational data, which is further analyzed/mined using ML
- ML results are used to generate data-driven hypothesis which drive experiments

Experimental data
generates a few features

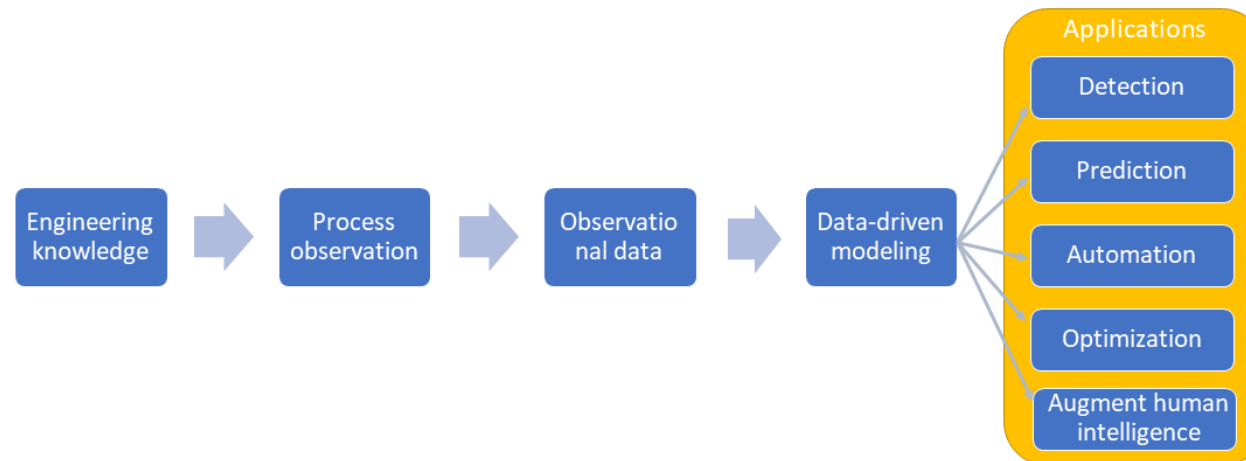


Observational data generates dozens, hundreds, or even thousands of features

Smart manufacturing

Data-driven modeling

- Enabled by the high data volumes, new modeling techniques, and increased computer power, data-driven modeling augments often even replace engineering and physics-based models
- Efficient approach to derive models from patterns and signals in the data itself
 - No assumptions about the form of the real-world model.
 - Allows engineers to automate a whole range of processes faster.
- Important to use engineering knowledge to avoid making the wrong decisions driven by wrong models



Smart manufacturing

Real-time iterations

- Real-time autonomous decisions, highest level of smart manufacturing
- Real-time data enables adjustments and actions in a timely manner to control the product quality
- Human-in-the-loop is always important

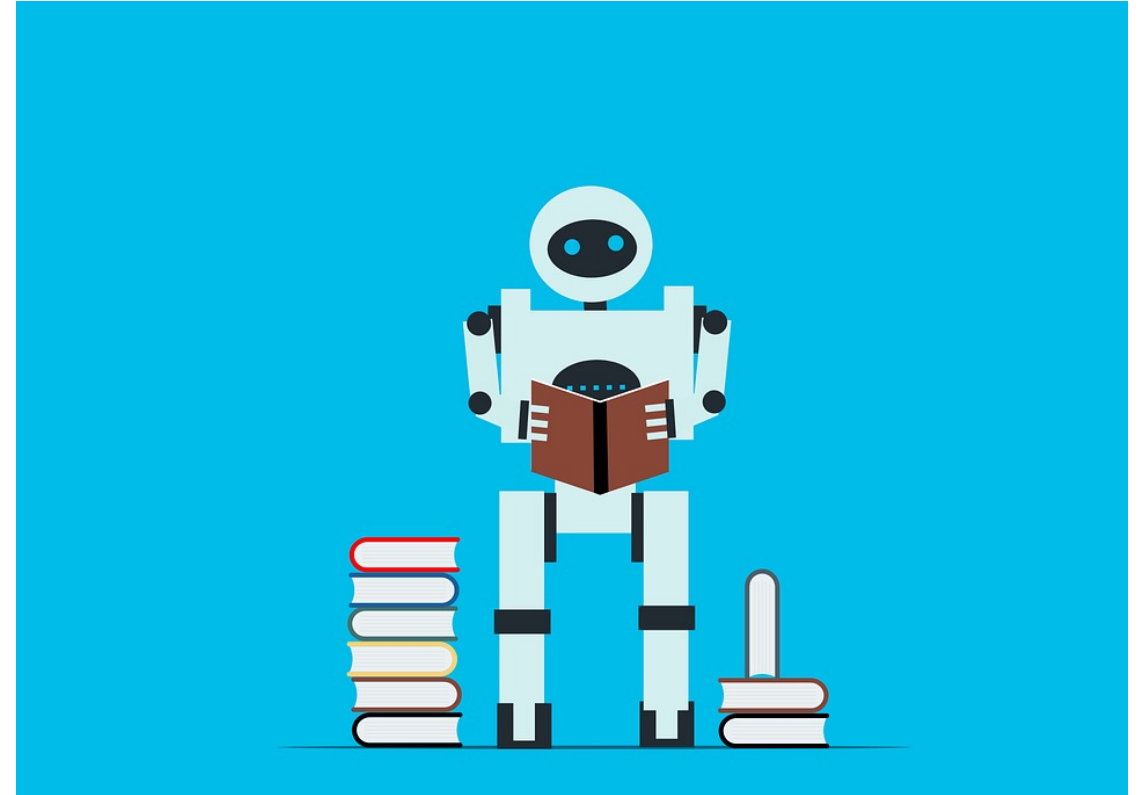


Source: <https://www.ictworks.org/help-usaid-find-evidence-of-real-time-data-use-in-development/#.Y1cmz3bMK38>

Smart manufacturing

Self learning adaptations and executions

- Manufacturing systems are constantly exposed to new sources of variations that generate new patterns
- Deployed system should have the capacity to automatically learn new patterns and automatically execute
- Self-learning code allows the system to keep improving without intensive human intervention as more data become available



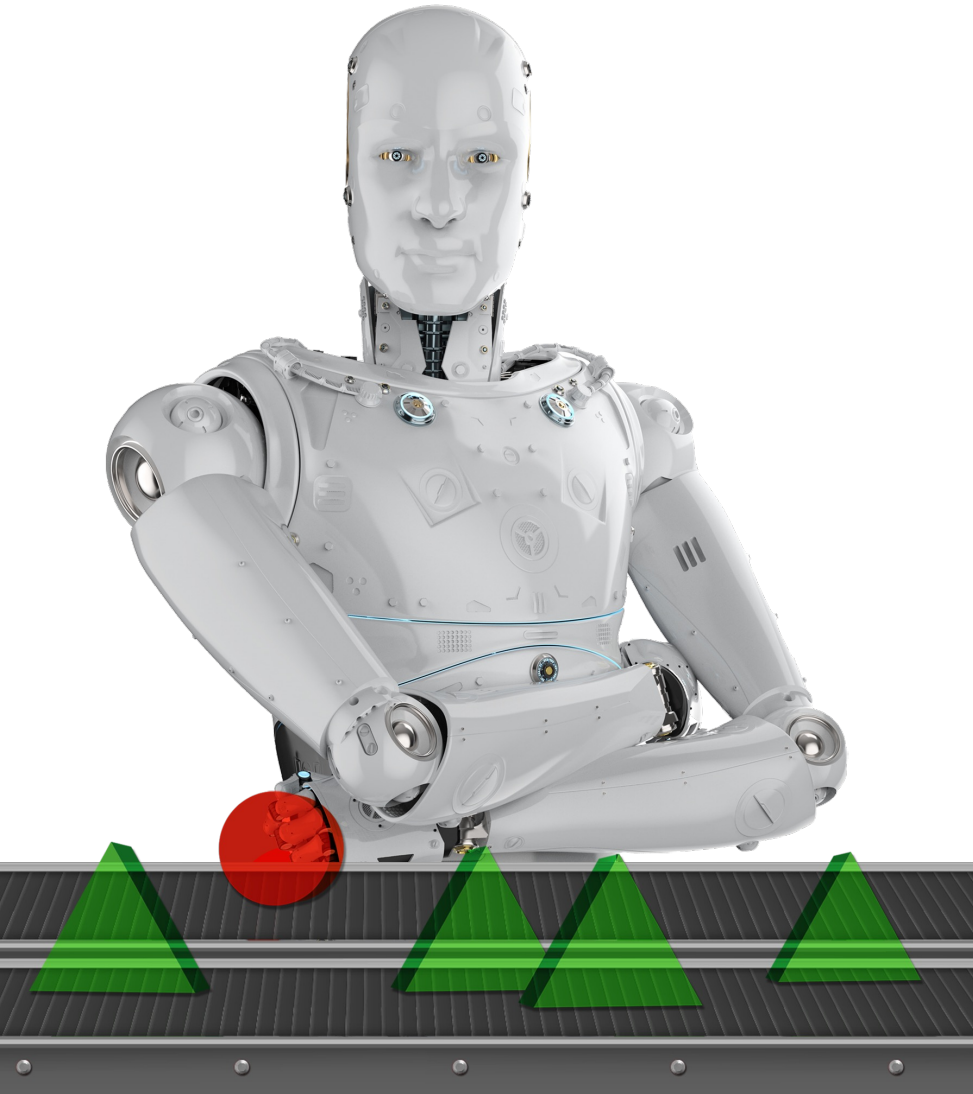
Source: <https://pixabay.com/vectors/robot-reading-machine-learning-6871146/>

Smart manufacturing

Remarks

- Systematic computational analysis of data enables the creation of intelligence in all the aspects of manufacturing
- Intelligence comes in the form of strategic actions, predictions, optimizations, simulations, and smart monitoring and control
- Most solutions should be executed in real-time, and self adapt to new sources of variations

Important paradigm in smart manufacturing



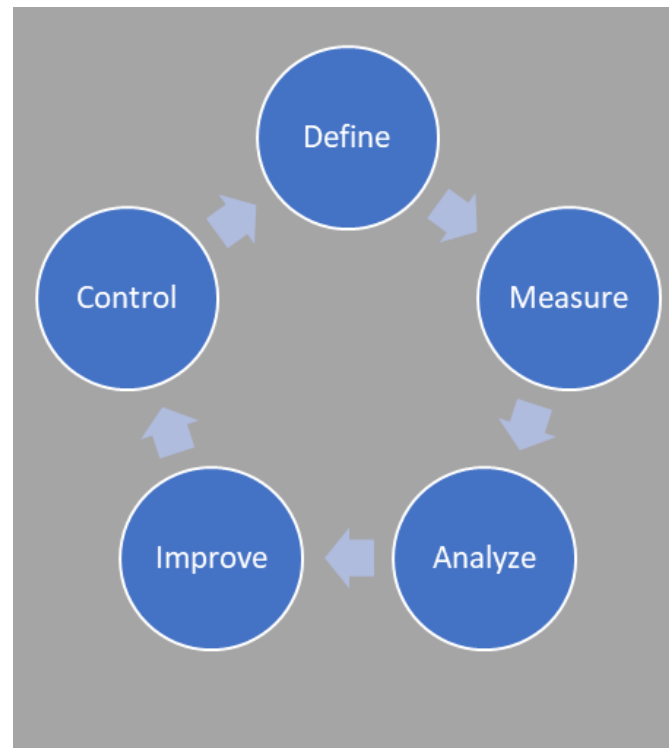
Modern quality movement

Modern quality movement

Problem solving strategy

- Iterative management method used in business for controlling and improving the quality of products and processes, resolving problems, discovering new engineering knowledge, and driving innovation.
- Adaptation of the scientific method to the modern QC

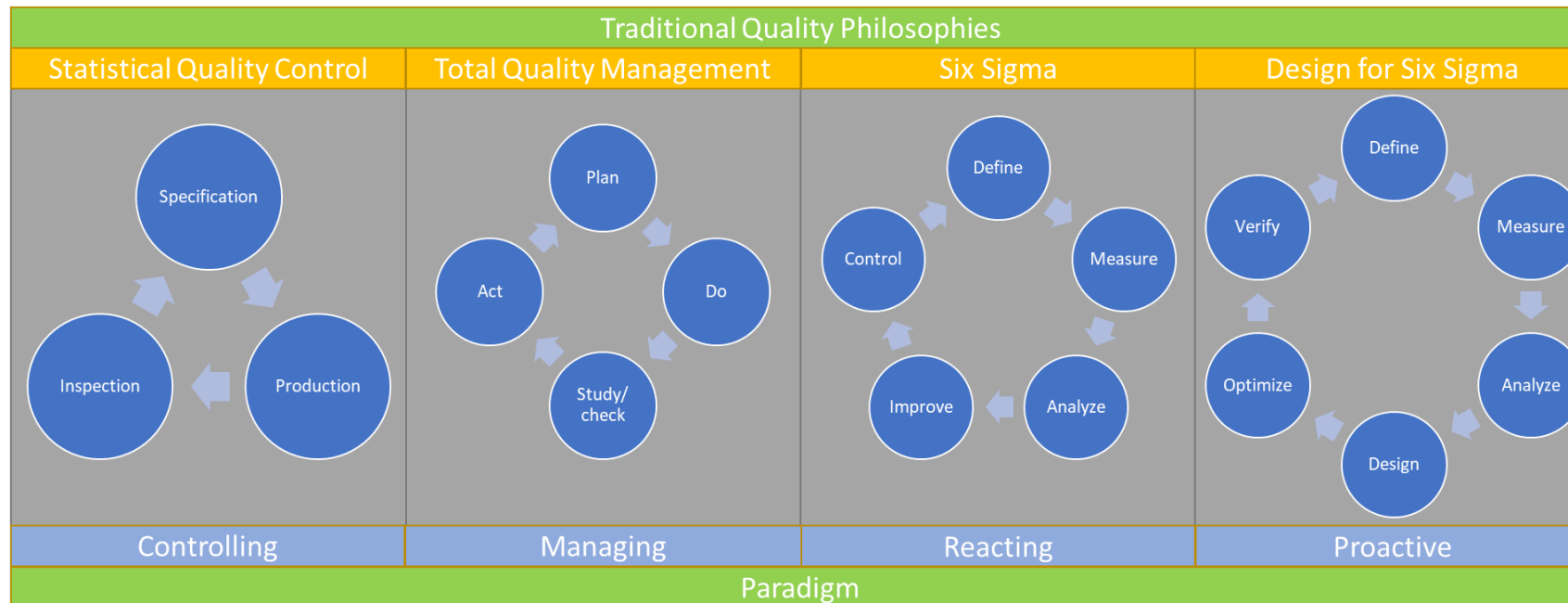
Six sigma DMAIC



Modern quality movement



Evolution of the manufacturing quality control movement

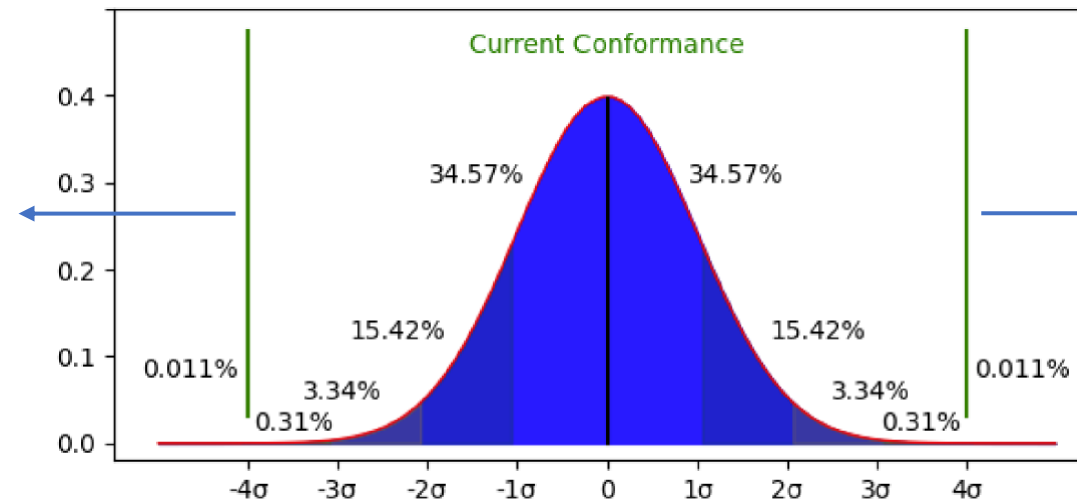


Modern quality movement

Plateau of six sigma and the rise of Quality 4.0

- Process monitoring charts and acceptance sampling methods have been widely implemented in manufacturing to set a process capability index benchmark of four Sigma
- Generating and integrating data from the whole value adding process and developing ML-based solutions enables to achieve the next sigma level (5σ , 233 DPMOs)

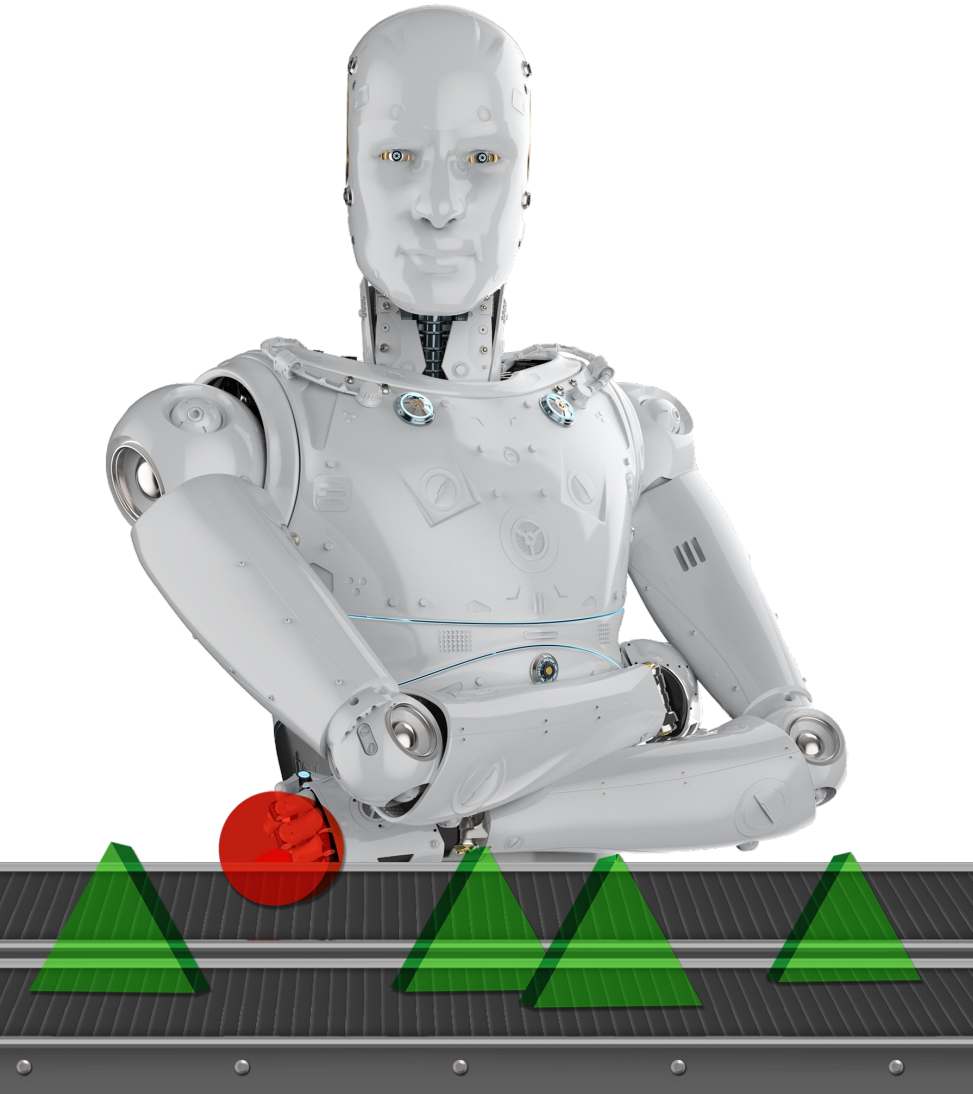
Challenge for Q4.0



Challenge for Q4.0

FIGURE 1.8 Four sigma conformance rate.

Breakdown of traditional quality control



Breakdown of traditional quality control

- Smart manufacturing is founded on processes that exhibit:
 - Rapidly increasing complexities
 - Hyperdimensional feature spaces
 - High data volumes
 - Transient variation sources
 - Reduced lifetime
 - Ultrahigh conformance
 - Non-Gaussian pseudo-chaotic behaviors
- Founding techniques and paradigms of traditional QC tools may not efficiently handle these dynamics



Breakdown of traditional quality control

Interest in Six Sigma

- Over the last decade quality profession stagnated with little innovation to offer to the manufacturing industry
- Quality professionals have lost their leadership positions to data scientists
- While traditional quality philosophies are still necessary, they are not driving innovations
- Interest in these have decreased

Traditional quality philosophies have plateaued, and they have stagnated with little innovation to offer to the manufacturing industry.

Source: Avigdor Zonnenshain and Ron S Kenett. Quality 4.0—the challenging future of quality engineering. Quality Engineering, 32(4):614–626, 2020.

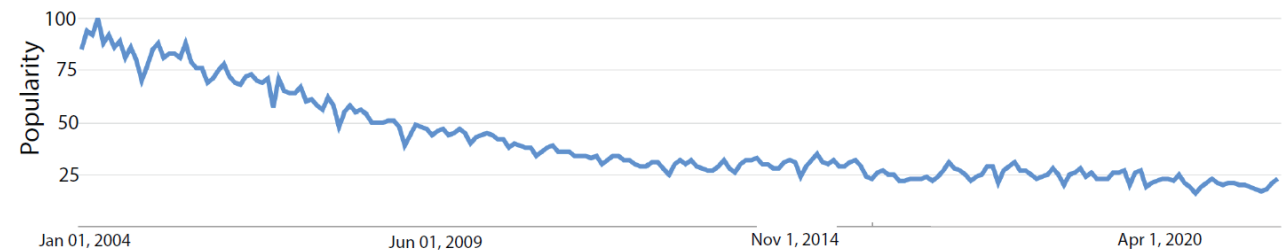
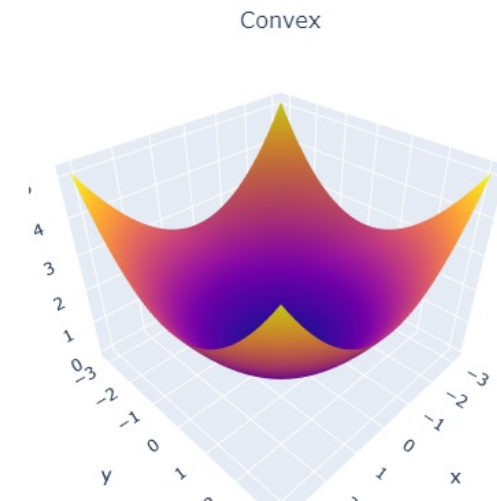
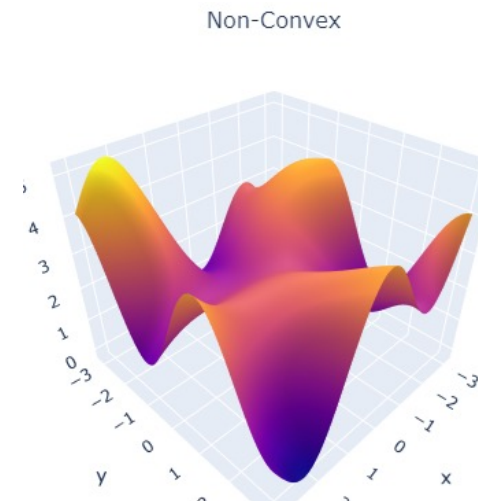
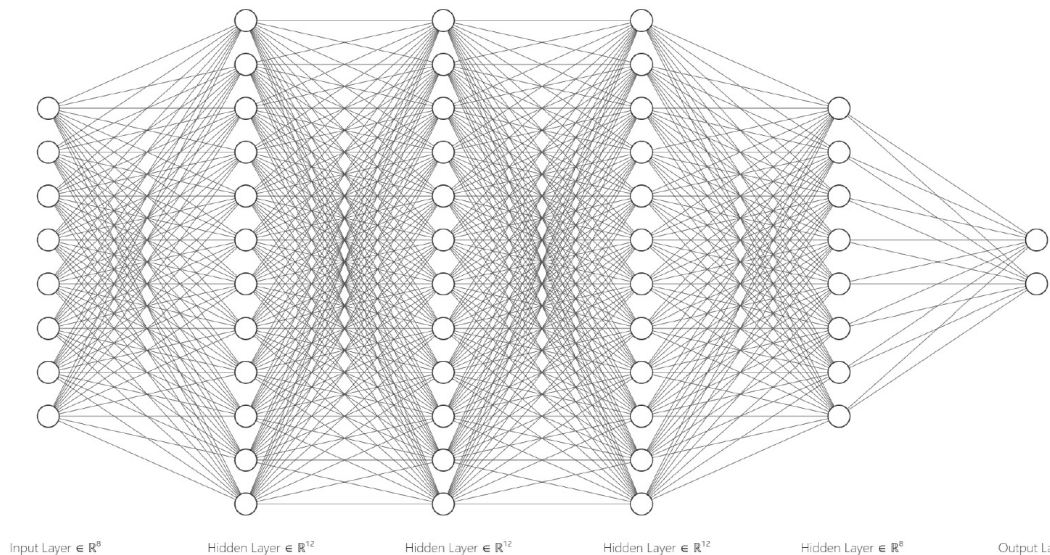


Figure 1. Google Trends, interest of Six Sigma from 2004 to 2022.

Breakdown of traditional quality control

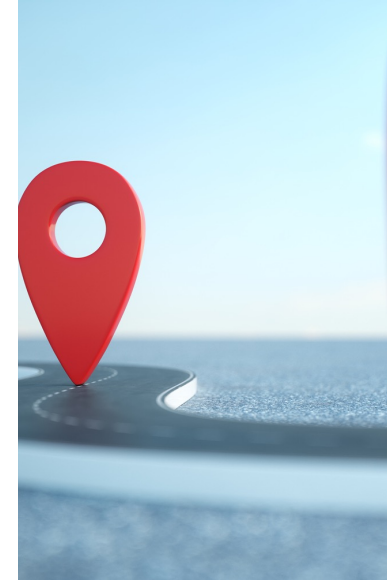
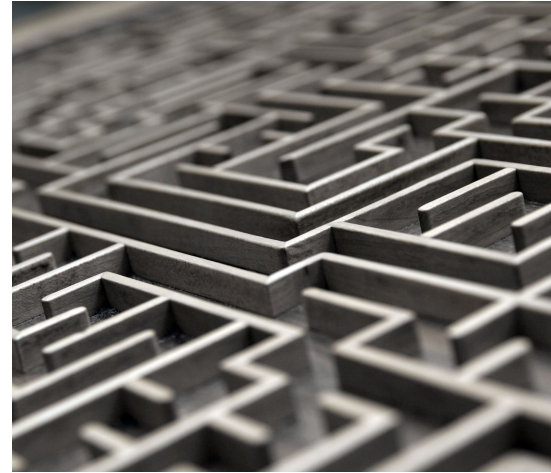
Limitations

1. ANN. Networks have shown better predictive modeling and optimization performance than response surface methodology (RSM). According to the universal approximation theorem, the feedforward ANN with one hidden layer containing sufficient neurons can approximate any continuous function to a reasonable accuracy level. This provides an advantage over the limitations of the quadratic models or cubic models used by RSM. Therefore, MLAs are better at learning complicated nonlinear relationships



Breakdown of traditional quality control

2. Curse of dimensionality. Traditional statistical modeling was designed for data having a few dozen input variables and sample sizes that would be considered small or moderate by the current standards. These methods can be highly affected by high-dimensional data. Whereas MLAs, can efficiently learn from high-dimensional data and non-linear patterns that exist in hyper-dimensional spaces



Breakdown of traditional quality control

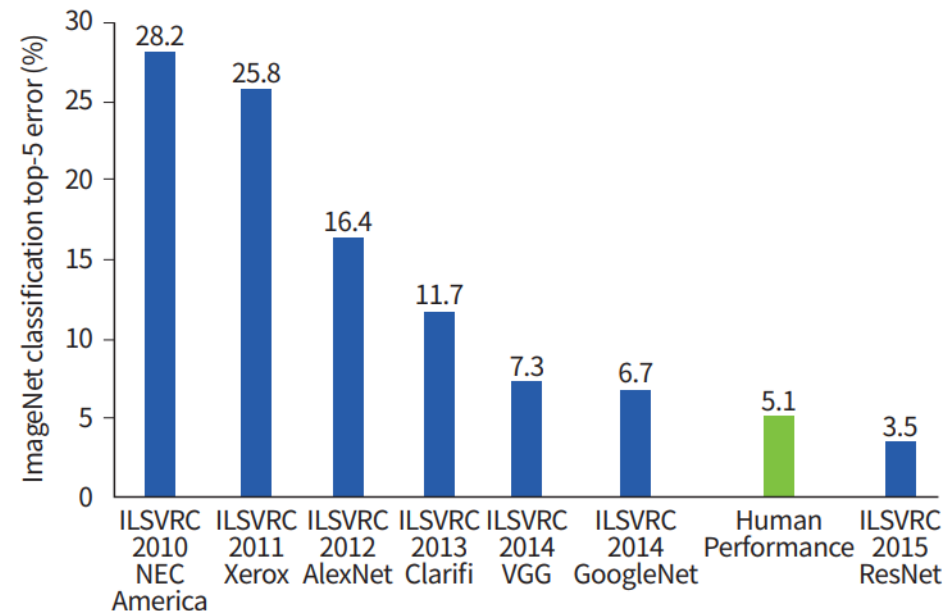
3. Computation time. There are two computational barriers for big data analysis: the data can be too big for a computer's memory, and the computing task can take too long to generate results. Most traditional statistical methods were not developed keeping in mind these challenges. However, modern statistical methods have been developed in view of these challenges. MLAs have embedded computational efficiency concepts to enable computational feasibility when learning from big data (e.g., stochastic gradient descent and XGBoost)



Source: <https://www.alltutorials.info/2018/01/how-to-auto-shutdown-computers-without-additional-software.html>

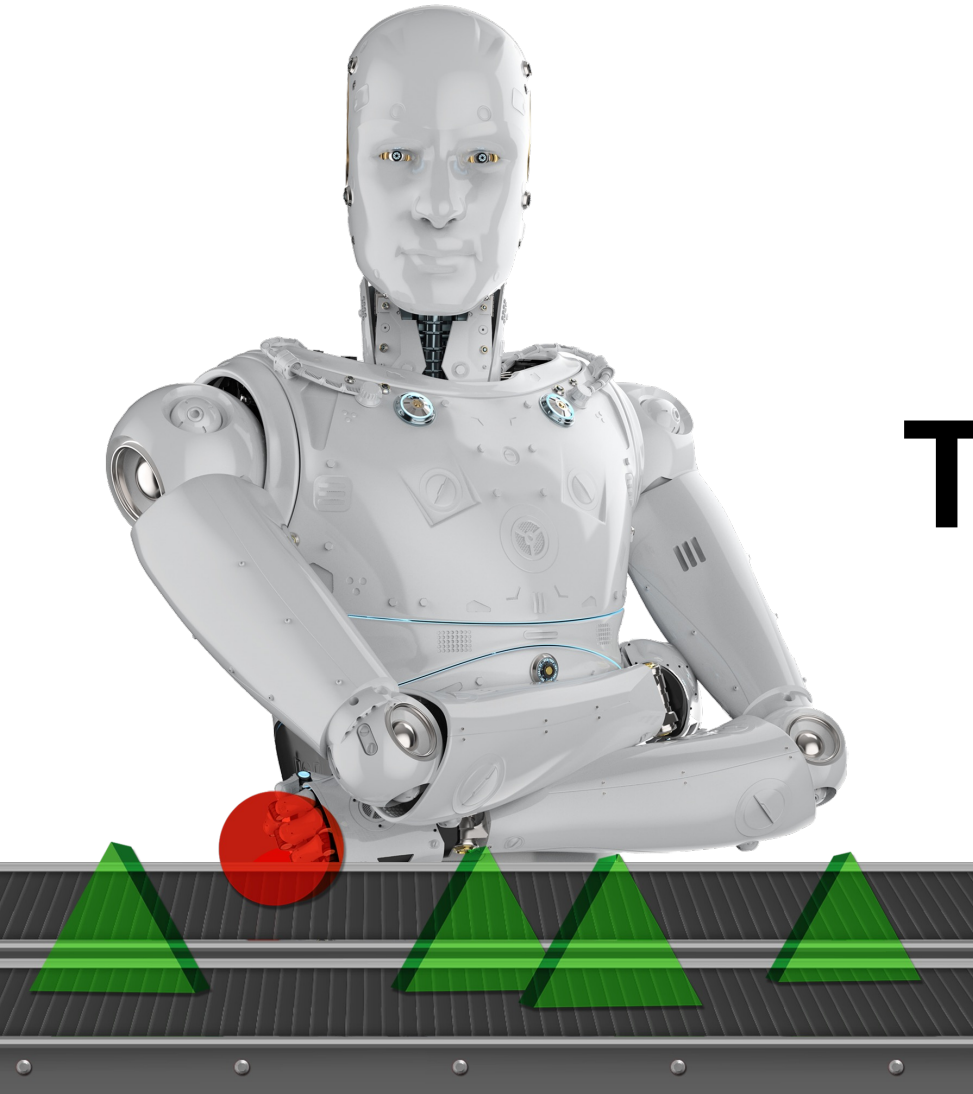
Breakdown of traditional quality control

4. Vision systems. Shortly after DL networks achieved superhuman performance on image and object recognition, a survey conducted in 2016 showed that almost half of the respondents claimed that their inspections were mostly manual. This was expected because DL was not part of the quality tools and had not achieved proper recognition performance when the traditional quality philosophies were developed. Manual inspections are often subject to the inherent operator biases (80% accuracy). Today, the development and application of vision systems for QC is one of the most relevant research topics in hyper-automation



Breakdown of traditional quality control

5. Control charts. These charts support fault detection and diagnosis of industrial processes and production results. However, univariate control charts cannot detect patterns in hyperdimensional spaces. Multivariate control charts, such as Hotelling's T-squared statistic and the Q-statistic are calculated based on a model using the Principal Component Analysis (PCA), which makes it difficult for this method to identify the process variables that lead to the quality problems. MLA can effectively learn patterns in hyperdimensional spaces and identify the driving variables of the process



The rise of quality 4.0

The rise of quality 4.0

Definitions

“the application of Industry 4.0’s advanced digital technologies to enhance traditional best practices in quality management”, *D. Küpper et al.*

“Quality 4.0 is the 4th wave in the modern quality movement. This quality philosophy is based on the statistical and managerial fundamentals of the previous philosophies. It leverages Manufacturing Big Data (MBD), IIoT, CSC, and AI to solve completely new sets of complex engineering problems. Quality 4.0 is based on a new paradigm that enables smart decisions through empirical learning, empirical knowledge discovery, and real-time data generation, collection, and analysis”, *Escobar et al.*

The rise of quality 4.0

“Quality professionals are going to have to master some of the skills of computer science, such as understanding the structure of large databases, basic data-mining techniques, image processing, and data visualization techniques”, *Montgomery*

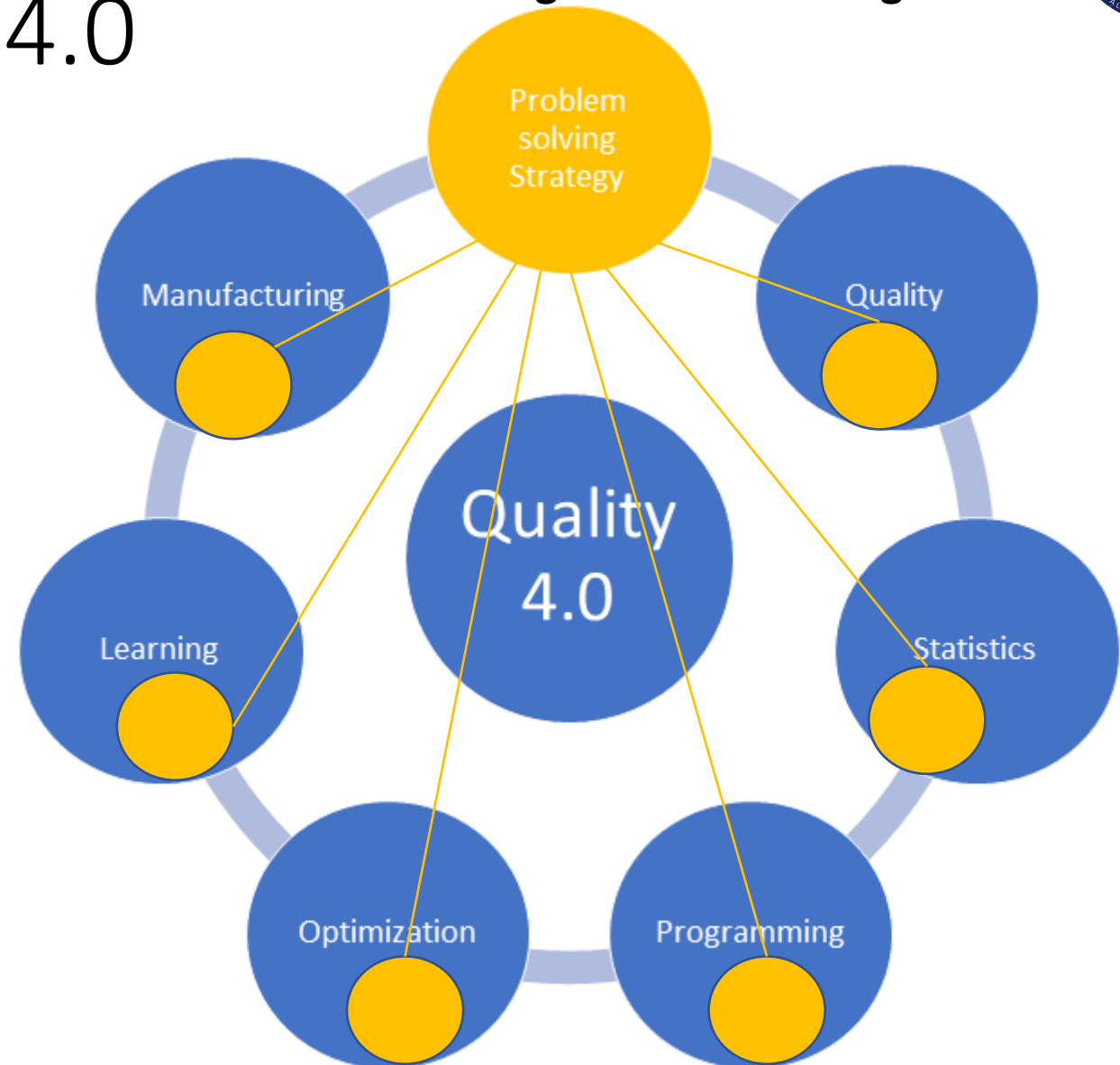
The rise of quality 4.0

- Quality 4.0 is the next natural step in the evolution of quality
- Intense modeling is the foundation of *Quality 4.0*
- Machine learning techniques are applied to the observational data generated by the manufacturing processes to develop empirical models and identify the driving features
- Traditional *QC* and *QA* philosophies are in the middle of a paradigmatic transformation that requires the full attention of management
- Fourth Industrial Revolution is an excellent opportunity for the quality movement to become relevant again and for quality professionals to return to lead roles again

The rise of quality 4.0

Introduce in the green belt training

Areas of knowledge





Quality 4.0

2020 survey findings

- 66% of the respondents believe that Quality 4.0 will significantly affect their operations within five years
- 16% have started to implement a Quality 4.0 initiative
- 20% respondents have started to plan for implementation
- 63% respondents have not even reached the planning stage yet
- zero respondents have done a full implementation

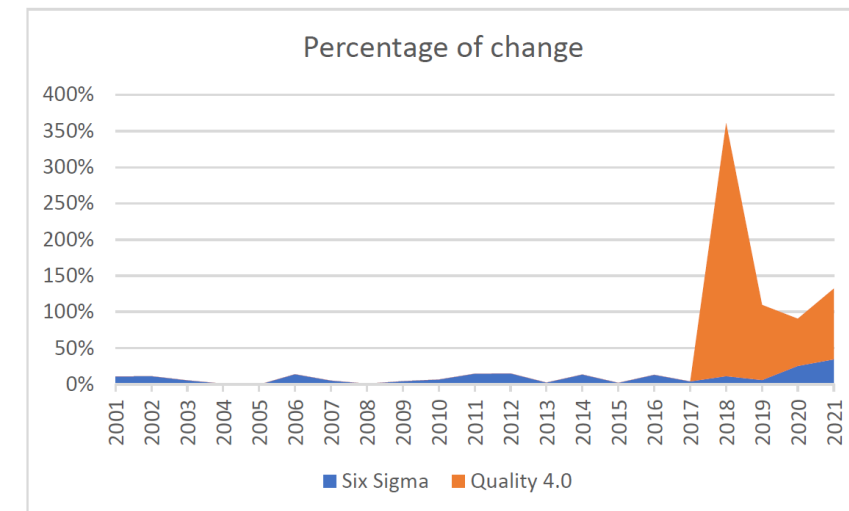
The rise of quality 4.0

Percentage of change of papers published in Quality 4.0 w.r.t. Six Sigma

- Quality 4.0 appeared in 2017
- Recent papers in Six Sigma are trying to frame AI, ML, and big data into the DMAIC frame

Manufacturing innovation highly rely on
Quality 4.0

Research community should be free
from Six Sigma and should stop trying to
fit the ML cycle into the DMAIC
problem-solving strategy



Source: Carlos A. Escobar, Daniela Macias-Arregoyta & Ruben Morales-Menendez (2023) The decay of Six Sigma and the rise of Quality 4.0 in manufacturing innovation, Quality Engineering, DOI: 10.1080/08982112.2023.2206679

The rise of quality 4.0

- Successfully implementing Quality 4.0 requires much more than just accessing the new technologies, generating data, and fitting a model
- It is necessary to understand and address many managerial and practical challenges
- Critical steps for successful implementation of Quality 4.0:
 - Develop a vision and road-map
 - Train the quality professionals in the new technologies
 - Allocate budget for the new technologies
 - Learn how to identify winning projects
 - Define a strategy for data generation
 - Understand the limitations of AI
 - Consider if outsourcing would save money and complete projects quickly, while in-house expertise is developed



Chapter 2: The Technologies

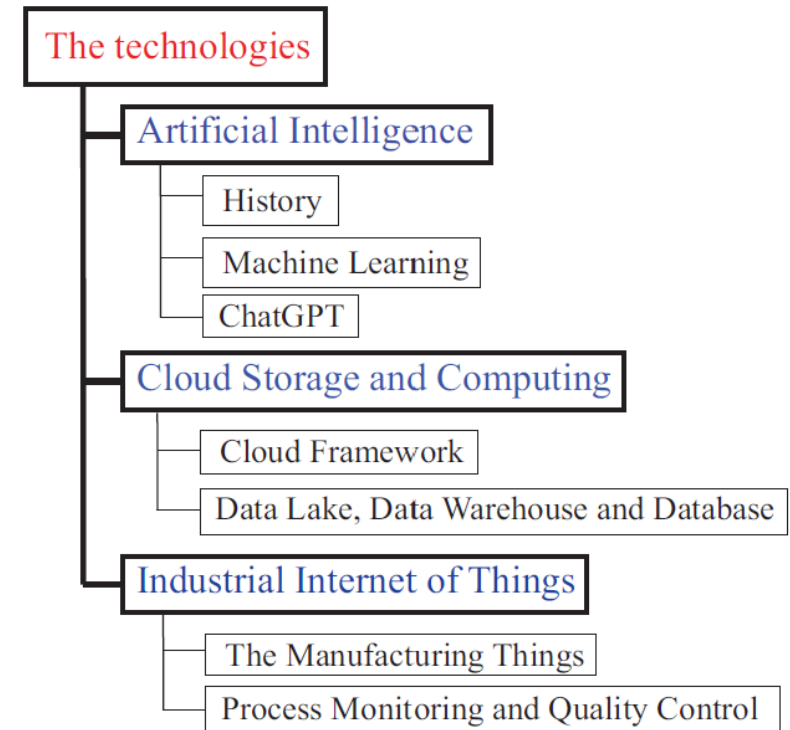
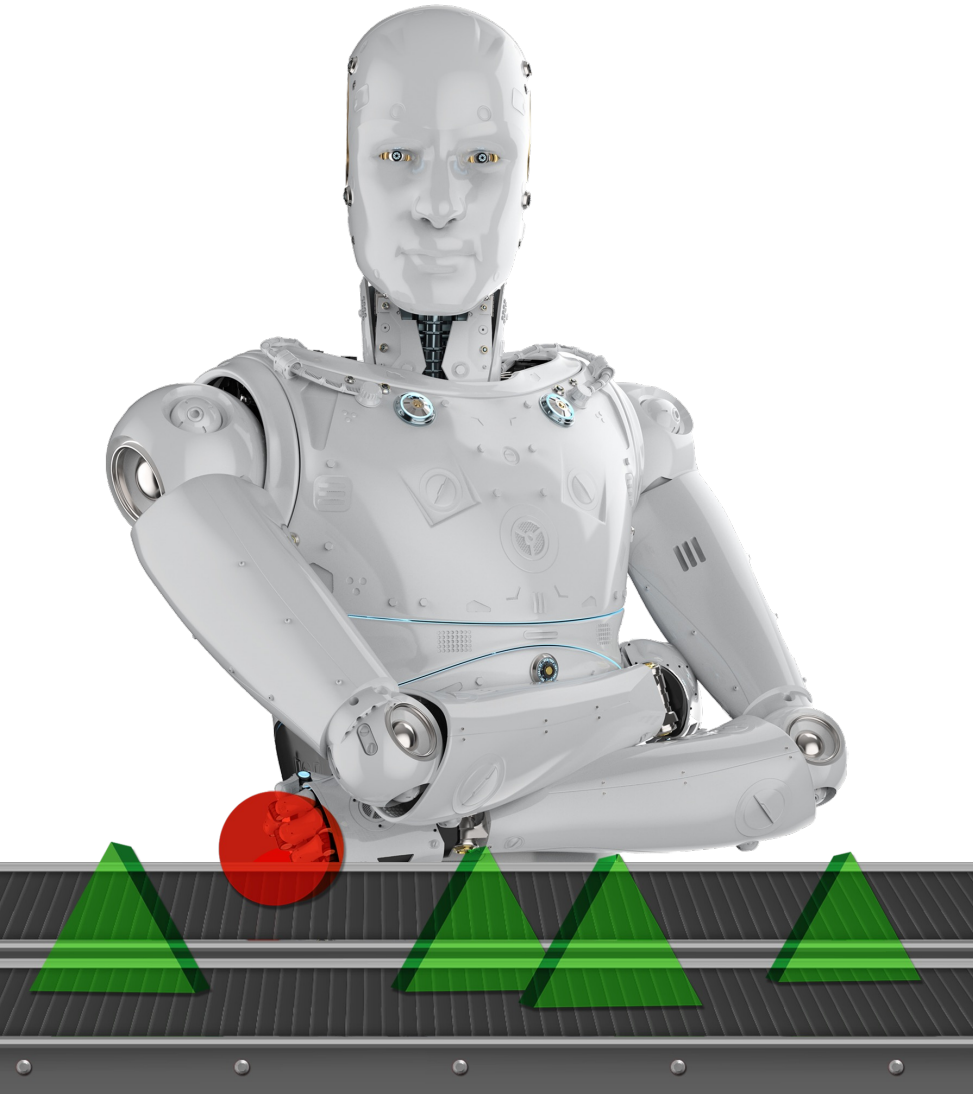


FIGURE 2.1 Contents of the Chapter 2



Artificial intelligence (AI)

Artificial intelligence

- Overall goal of AI is to create technologies that understand, think, learn, and behave like humans
- These technologies have the capacity to augment human intelligence, replace monotonous tasks, and speed up scientific discoveries
- They are programmed to autonomously handle situations with little or without human intervention

Narrow AI



Source: <https://askleo.com/email-lot-bcced-recipients-bounce/>

General AI



General AI warning



Source: <https://truththeory.com/stephen-hawking-developing-spacecraft-capable-reaching-second-earth-20-years/>

Artificial intelligence

- Although the history of AI dates so many decades far back, the convergence of big data, CSC and IIoT in the last decade, allowed scientist and engineers to build real-life applications
- Today worldwide leaders consider AI the most disruptive technology
- Experts believe AI automation will replace 40% of jobs in 15 years

Artificial intelligence

Statistics

- Latest surveys estimated that the market size of this technology was valued at \$51.08 billion in 2020
- Projected to reach \$641.3 billion by 2028
- More than tenfold increase in just eight years and with a Compound Annual Growth Rate (CAGR) of 36.1%
- By 2030, roughly 70 % of the companies may have adopted at least one type of AI technology
- Generating additional \$13 trillion in economic growth
- Only 17 % of senior decision-makers in the US are familiar with AI
- Although, they understand the potential of this technology, they lack a clear deployment strategy

Artificial intelligence

Machine learning

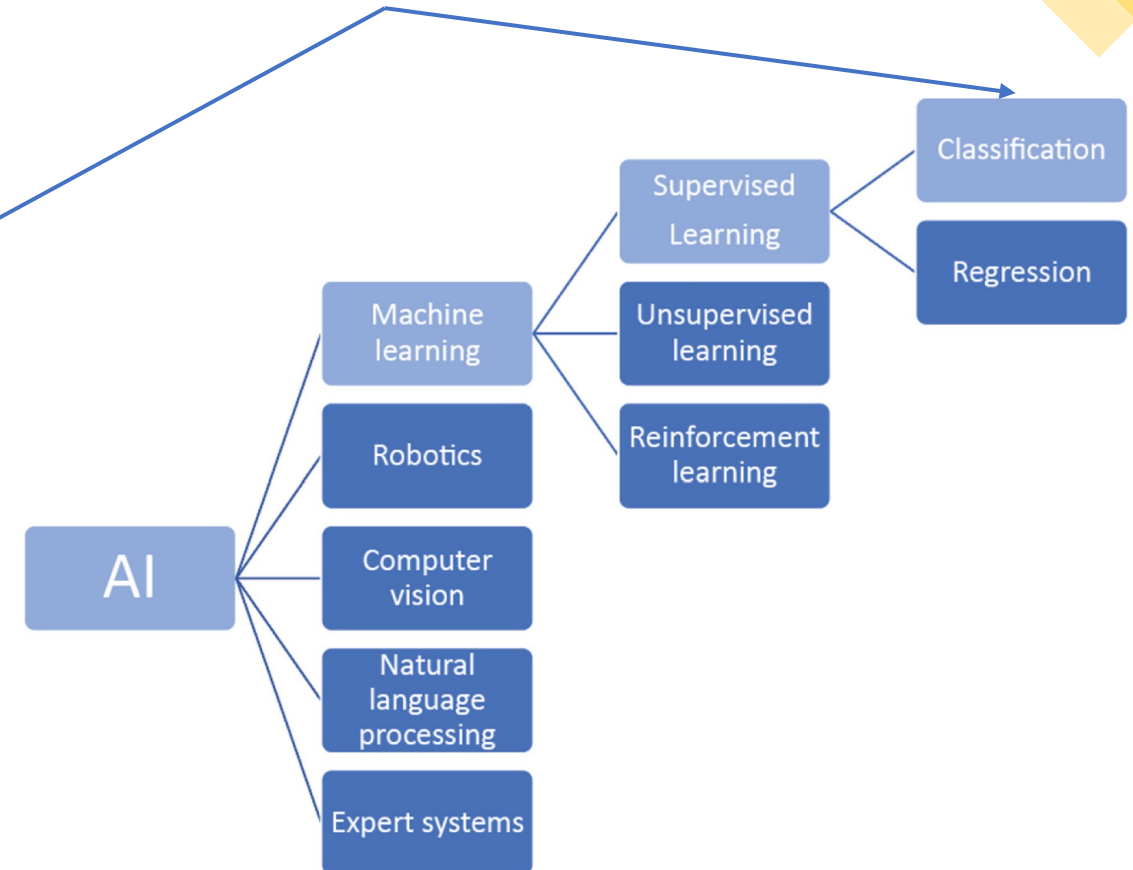
- Subfield of AI that gives computers the ability to learn from examples without explicitly being programmed
- MLAs play a crucial role in the structure of AI.
- They become smarter with more data
- Most MLAs automatically learn patterns without assuming a probabilistic distribution or a predefined model form
- Simple algorithms are used in simple applications, while more complex ones help to solve complicated engineering problems

Artificial intelligence

Foundation of LQC

ML techniques are classified into three categories

- **Supervised learning**
 - All observations in the dataset are labeled, and the MLA learns to predict the output using the input data
 - Supervised models are further categorized into regression and classification methods
- **Unsupervised learning.**
 - All observations in the dataset are unlabeled, and the MLA learns its inherent structure using the input data
 - Clustering, dimensionality reduction, density estimation, and association methods are included in this category
- **Reinforcement learning**
 - Family of models consists of an MLA, which uses the estimated errors as rewards or penalties. The goal is to identify the best actions that maximize the long-term reward
 - Trial-and-error searching, and delayed reward are the most essential characteristics of this approach



Artificial intelligence

Supervised learning

- All the observations in the data set are labeled
- MLA learn to predict the output using the input data (features)
- Regression
 - Predict house value – continuous target value

Sample	Feature 1	Feature 2	Feature 3	.	.	.	Feature 45	Value
1	-0.57	-0.88	-1	-0.86	-0.51	-1.01	-1.03	0.56
2	0.17	0.17	0.25	0.03	0.67	0.39	0.09	1.88
3	1.11	1.26	1.1	1.31	1.42	1.17	1.15	0.67
.	0.24	0.18	0.3	0.36	0.32	0.55	0.22	0.45
.	-1.56	-1.15	-0.33	-1.52	-1.34	-1.27	-1.24	1.15
.	-1.73	-1.56	-1.47	-1.52	-1.55	-1.44	-1.37	0.29
421	-0.41	-0.5	-0.52	-0.14	-0.46	-0.33	-0.23	1.14



Source: <https://medium.com/analytics-vidhya/predicting-house-price-with-random-forest-regressor-6e9a7c7f75dd>

Artificial intelligence

Classification

- Diagnose cancer or detect a defective item
– categorical target value

Sample	Feature 1	Feature 2	Feature 3	.	.	.	Feature 155	Label
1	−0.57	−0.88	−1	−0.86	−0.51	−1.01	−1.03	0
2	0.11	0.17	0.25	0.03	0.67	0.39	0.49	1
3	1.11	1.16	1.1	1.31	1.42	1.17	1.15	0
.	0.21	0.18	0.3	0.31	0.32	0.55	0.22	0
.	−1.5	−1.15	−0.33	−1.52	−1.34	−1.27	−1.14	1
.	−13	−1.56	−1.47	−1.52	−1.5	−1.44	−1.37	1
1,221	−0.41	−0.5	−0.51	−0.14	−0.44	−0.33	−0.23	1

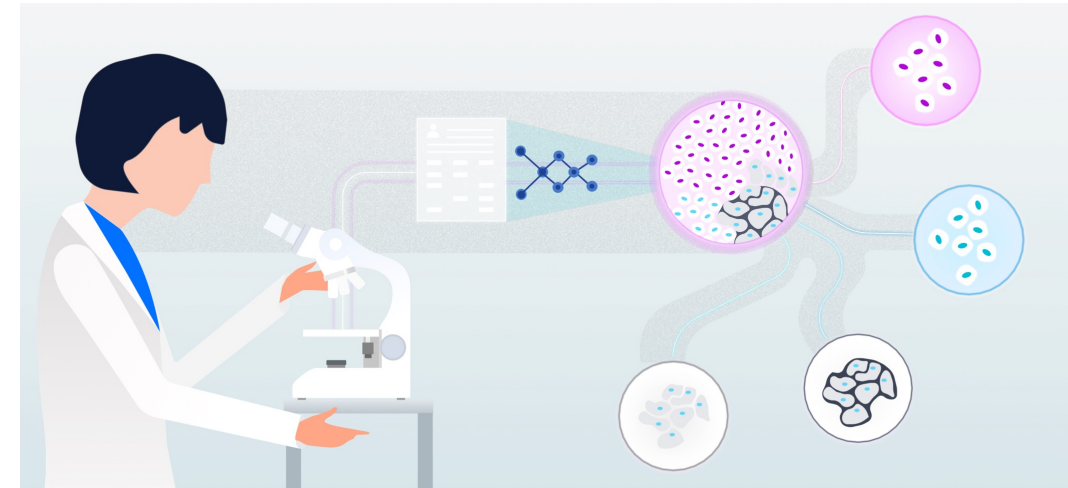


Image source: <https://github.com/gscdit/Breast-Cancer-Detection>

Artificial intelligence

Desirable characteristics of machine learning projects

- Learning a simple concept with a lot of data is more likely to be feasible
- An activity that requires only a few number of seconds of mental processing
- Quick wins are good starting points
- Project that can yield values within 6-12 months
- Once the data science teams have gained traction, more complex problems can be pursued
- Projects should be relevant to the industry and should leverage the know-how of the company
- Projects should address the impossible vs straightforward as well as the ambiguous vs specific trade-offs
- Should not be impossible but should also not address simple problems that can be solved using conventional methods (e.g., t-test)

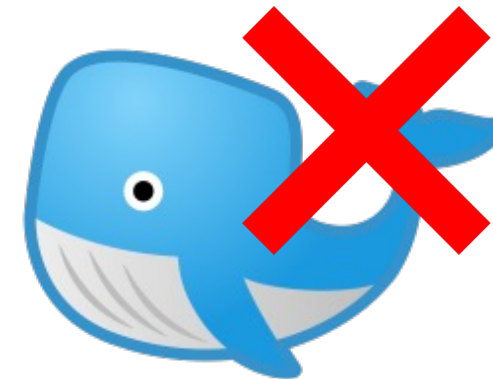
Artificial intelligence

Many projects are ill-conditioned from the start

- Avoid complex applications with little data available
- Avoid to model complex and chaotic systems
 - Chasing a whale
 - Bitcoin value



My cousin Alex and I chasing whales in Cabo



Artificial intelligence

Five points that entrepreneurs interested in integrating AI into their business models should keep in mind during the first 90 days:

- Identify the problem or opportunity that AI can help with
- Understand the capabilities and limitations of AI
- Consider the potential risks and ethical implications
- Plan for the integration and adoption process
- Monitor and evaluate the results

Source: <https://www.inc.com>

We Asked ChatGPT and Sophia the Robot to Predict the Impact of A.I. on the Business World. Here's What They Said, By Ben Sherry

Sophia speaking at the AI for GOOD Global Summit, International Telecommunication Union, Geneva, June 2017
Image source: <https://www.smithsonianmag.com/smart-news/saudi-arabia-gives-robot-citizenship-and-more-freedoms-human-women-180967007/>

Artificial intelligence



ChatGPT

Artificial intelligence chatbot developed by OpenAI and launched in November 2022.

- Based on Reinforcement Learning from Human Feedback (RLHF)
- Language model trained on a massive amount of text data from the internet
- Designed to understand and generate human-like language
- Available for anyone to try out for free
- Answer questions, converse on a variety of topics, write and debug code, and generate creative writing piece

Source: <https://openai.com/blog/chatgpt>

Activity 2

Access the following link <https://openai.com/blog/chatgpt> and make two requests, 10mins

“write three paragraphs about quality 4.0”

“write the python code to compute the Pearson correlation coefficient between two random variables”

“what is the definition of the Pearsons correlation coefficient”

Discussions, 5mins

Activity 2



write three paragraphs about quality 4.0

Results



Quality 4.0 is a term used to describe the integration of Industry 4.0 technologies with quality management practices. The aim of Quality 4.0 is to enhance product quality and process efficiency through the use of big data analytics, machine learning, and other advanced technologies. Quality 4.0 represents a paradigm shift in quality management, moving from a reactive, inspection-based approach to a proactive, data-driven approach.



Quality 4.0 has several key components, including the use of sensors and other data collection devices to monitor production processes in real-time, the integration of data from different sources to gain a holistic view of the production process, and the use of advanced analytics to identify patterns and trends in the data. These components allow manufacturers to make data-driven decisions about process optimization, quality control, and product design, which can lead to significant improvements in product quality and production efficiency.

One of the main benefits of Quality 4.0 is that it allows manufacturers to move from a reactive, inspection-based approach to quality management to a more proactive, data-driven approach. By continuously monitoring and analyzing production data, manufacturers can identify potential quality issues before they become major problems, and take action to prevent defects from occurring. This can lead to significant cost savings, as well as improvements in customer satisfaction and brand reputation. Overall, Quality 4.0 represents a major step forward in the evolution of quality management, and has the potential to transform the manufacturing industry.

Activity 2



write the python code to compute the pearson correlation coefficient between two random variables



You can use the `pearsonr` function from the `scipy.stats` module to compute the Pearson correlation coefficient between two random variables. Here's an example:



```
python Copy code

import scipy.stats as stats
import numpy as np

# Generate two random variables
x = np.random.rand(100)
y = np.random.rand(100)

# Compute the Pearson correlation coefficient
r, p = stats.pearsonr(x, y)

print("Pearson correlation coefficient:", r)
```

In this example, we first generate two random variables `x` and `y` using the `rand` function from the `numpy` module. Then we use the `pearsonr` function from the `scipy.stats` module to compute the Pearson correlation coefficient between `x` and `y`. The function returns two values: the correlation coefficient `r` and the p-value `p` (which measures the significance of the correlation). We print out the correlation coefficient `r` using the `print` function.



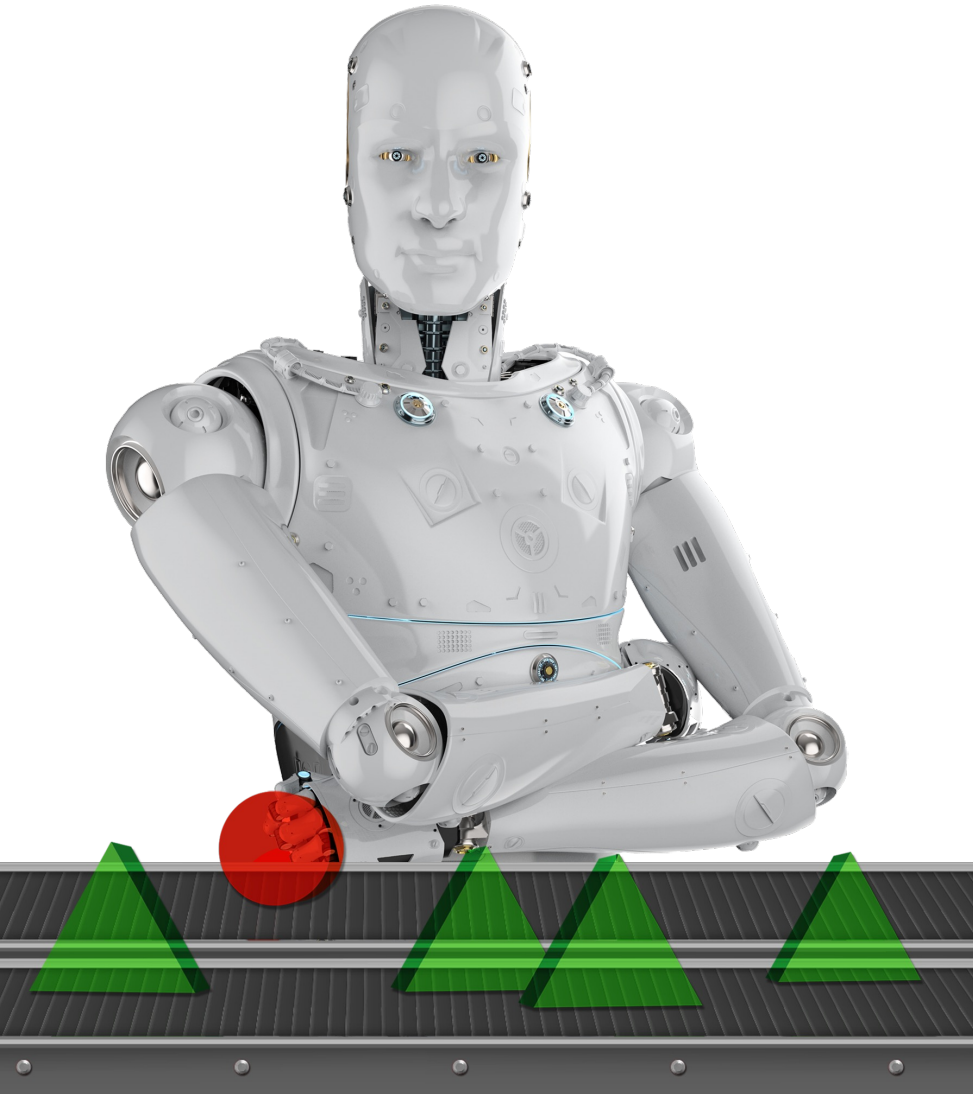
what is the definition of the Pearsons correlation coefficient



The Pearson correlation coefficient is a statistical measure of the strength and direction of the linear relationship between two continuous variables. It is represented by the symbol "r" and ranges from -1 to +1. A value of +1 indicates a perfect positive linear relationship, where the two variables increase or decrease together at the same rate. A value of -1 indicates a perfect negative linear relationship, where the two variables move in opposite directions at the same rate. A value of 0 indicates no linear relationship between the two variables.



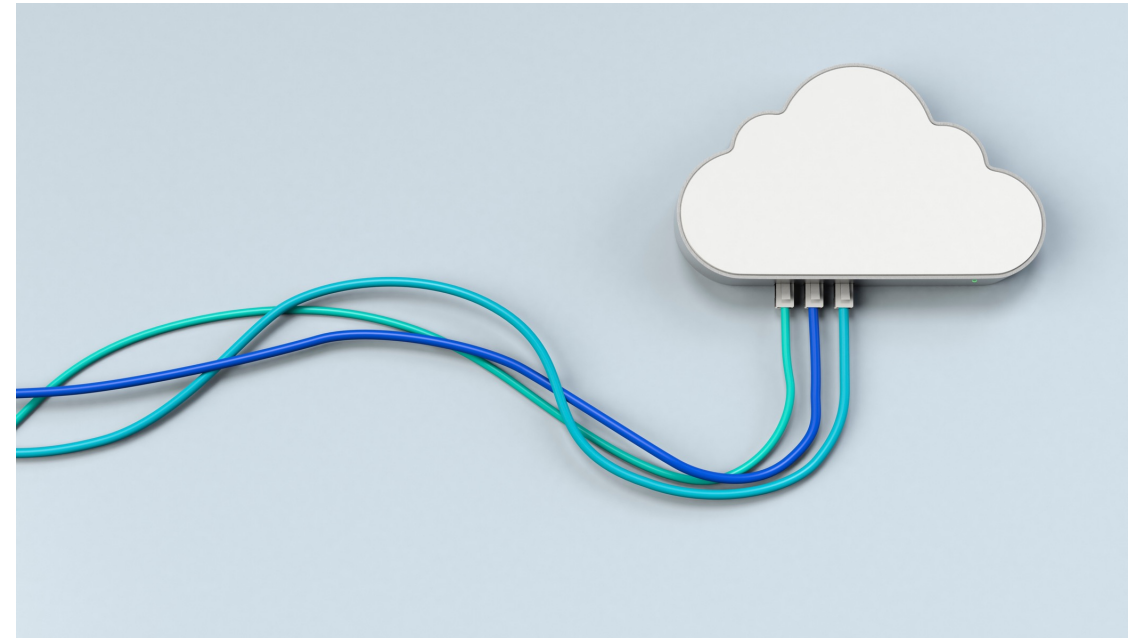
The Pearson correlation coefficient is calculated by dividing the covariance of the two variables by the product of their standard deviations. It is commonly used in data analysis and research to determine the degree of association between two variables, and to test hypotheses about the nature of this association. However, it should be noted that the Pearson correlation coefficient only measures the strength of linear relationships and is not appropriate for non-linear relationships or relationships between categorical variables.



Cloud storage and computing (CSC)

Cloud storage and computing

- MBD is constantly increasing in volume, diversity, and complexity
- Innovation requires cost-effective ways of rapid information processing
- Traditional on-premises storage and computing systems may not be able to cope with this challenge
- Cloud Storage and Computing (CSC) accelerates innovation, engineering and scientific discovery
- With access to the internet, companies can immediately leverage high computing power to train complex ML models
- Allowing them to focus on what really matters for them, avoiding undifferentiated work like procurement, maintenance, and capacity planning



Cloud storage and computing

Statistics

- 75% of the cloud's predicted value comes from boosting innovation
- Accelerate discoveries up to 10 times faster and with 10 times less cost
- Worldwide end-user spending on public cloud services to grow year-over-year 20.4% in 2022 to total \$494.7 billion.
- In 2023, end-user spending is expected to reach nearly \$600 billion

Cloud storage and computing

Benefits

- Flexibility and scalability
- Faster deployments
- Democratization of AI
- Simple and cost-effective
- Accelerated discovery

Cloud storage and computing

Cloud Framework for Quality 4.0

- **Offline learning**
- Big data analyses conducted to train the deployable model, classifier (f)
- Historical process data and apply ML techniques (supervised learning) to learn the quality patterns.
 - Features (X)
 - Quality label (y)
- MLAs learn the coefficients θ of the general form $\hat{y} = f(X; \theta)$
- All the analytical work required to develop the LQC system is performed here
- Activity can last for a couple of weeks or several months

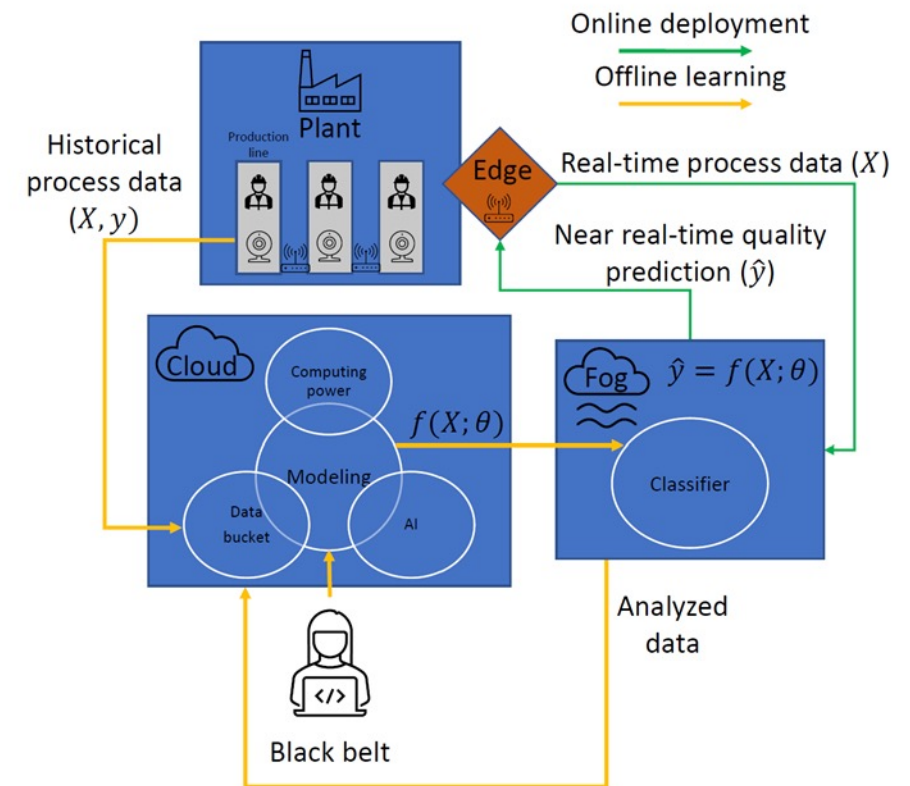


FIGURE 2.3 Data and technologies scheme supporting the offline learning and online deployment of the LQC system.

Cloud storage and computing

Cloud Framework for Quality 4.0

- **Online deployment**
- Classifier into production for quality control
- Real-time process data (X) is run through the previously trained classifier (f) to predict, in near real-time, the probability of each item to belong to the defective class ($\hat{y}$)
- Online deployment requires low latency to support real-time or near real-time responses

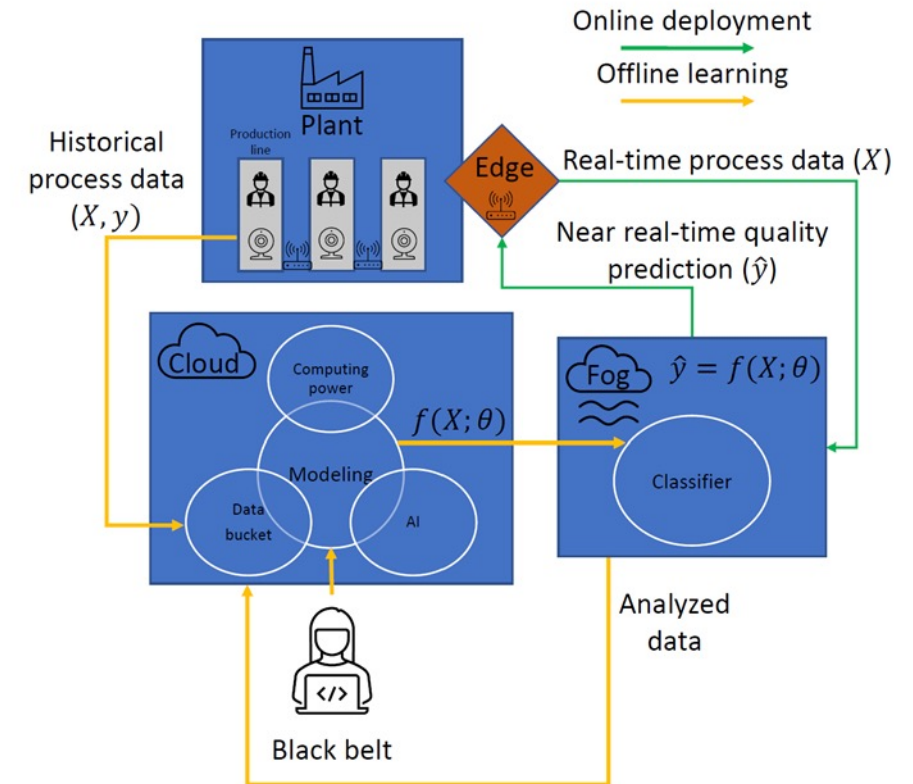
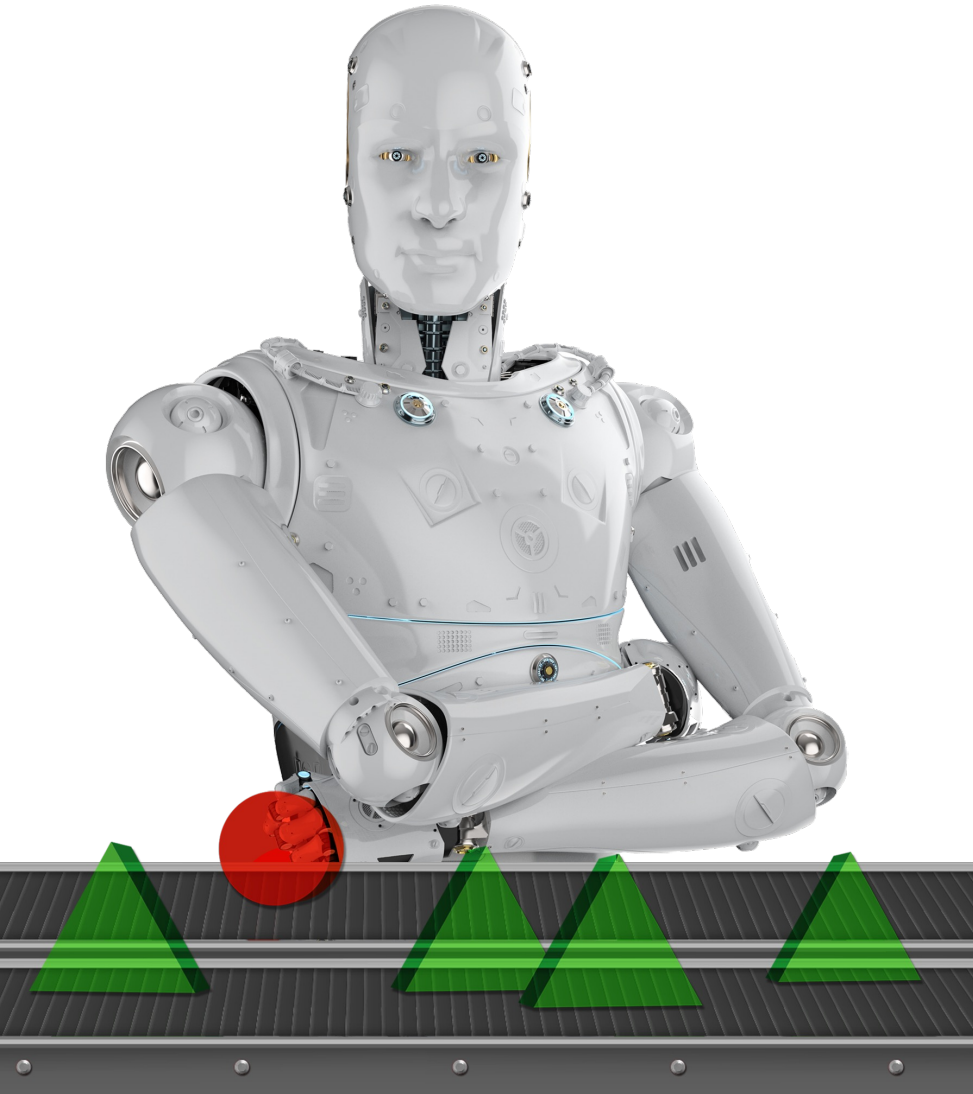


FIGURE 2.3 Data and technologies scheme supporting the offline learning and online deployment of the LQC system.



Industrial internet of things (IIoT)

Industrial internet of things

Definitions

...the network of physical objects-“things”-that are embedded with sensors, software, and other technologies for the purpose of connecting and exchanging data with other devices and systems over the internet. *Oracle*

...a physical element that can be identified throughout its life and interact with the environment and other objects. Moreover, it can act in an intelligent way and independently under certain conditions. Furthermore, Smart Objects have an embedded operating system, and they usually can have actuators, sensors, or both. *Cristian González García, Daniel Meana-Llorián, Juan Manuel Cueva Lovelle, et al.*

Amazon's voice assistant Alexa, is one of the most illustrative examples of IoT services

Statistics

In 2025, there will be approximately 27 billion connected devices in the world.

Industrial internet of things

- IIoT technology refers to the application of IoT in industry settings.
- In the manufacturing industry, IIoT implies:
 - Machine-to-machine communications
 - Use of sensors and actuators to develop process monitoring and control systems founded on data analytics.

Industrial internet of things

The Manufacturing Things

- Sensors monitor conditions of equipment and processes, whereas actuators drive events based on what is sensed
- Sensors and actuators are often found in the same areas of equipment and processes
- Although they often interact, they are two different components that complement each other in developing intelligent applications

Industrial internet of things

Smart sensors serve as the interface between the physical and virtual worlds. A smart sensor is a device that combines a sensor, microprocessor, and communication technology to convert environmental inputs into readable data, that is transmitted to a centralized computer system for real-time analysis.

- Optical image sensors
- Position sensors
- Pressure sensors
- Temperature sensors
- Force sensors
- Fluid property sensors
- Humidity sensors
- Microphones
- Photo optic sensors
- Linear variable differential transformer
- Accelerometers

Industrial internet of things

Actuators

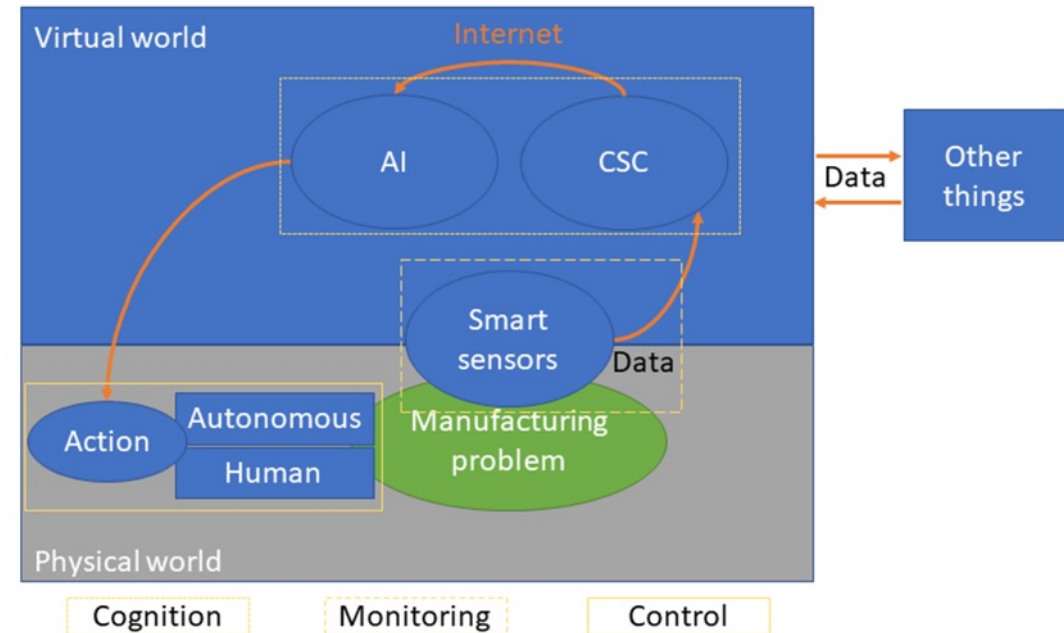
An actuator can be a mechanical device that makes something moves or operates.

- Common actuators in industry are:
 - Electric motors
 - Hydraulic bombs
 - Pneumatic control valves.

Industrial internet of things

Process Monitoring and Quality Control in Manufacturing

- IIoT enable the creation of CPS by connecting the manufacturing “things” into a network that leverages CSC and AI technologies to solve manufacturing problems and drive innovation
- Production data is generated, collected, shared, and analyzed in real-time to take actions or smart decisions
- Decisions/actions may be taken by humans or autonomous machines

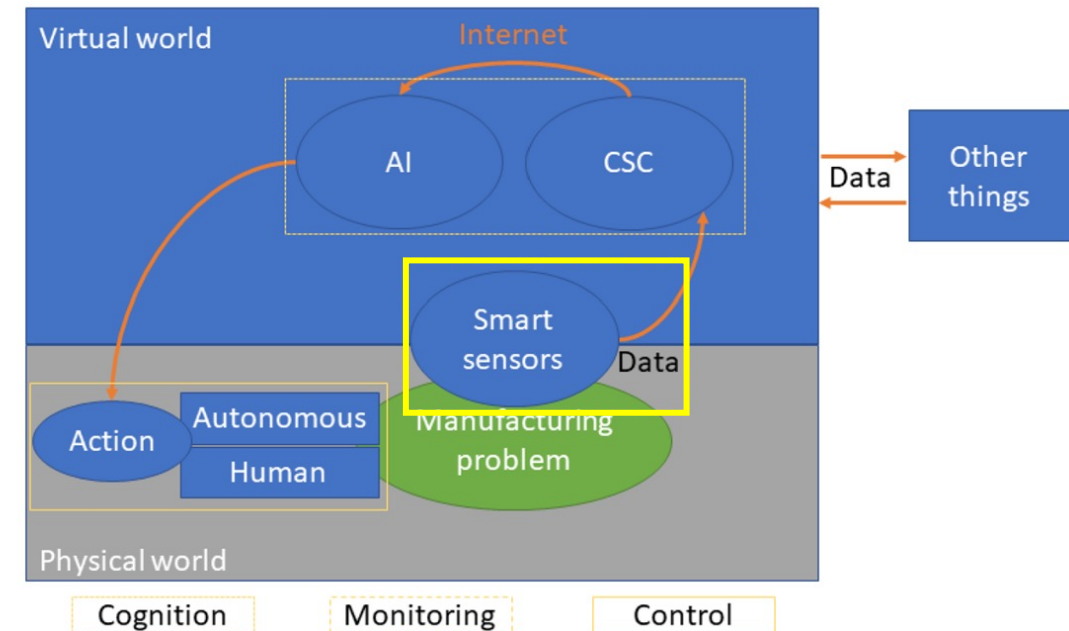


Industrial internet of things

CPS in manufacturing for quality control have three components:

Monitoring

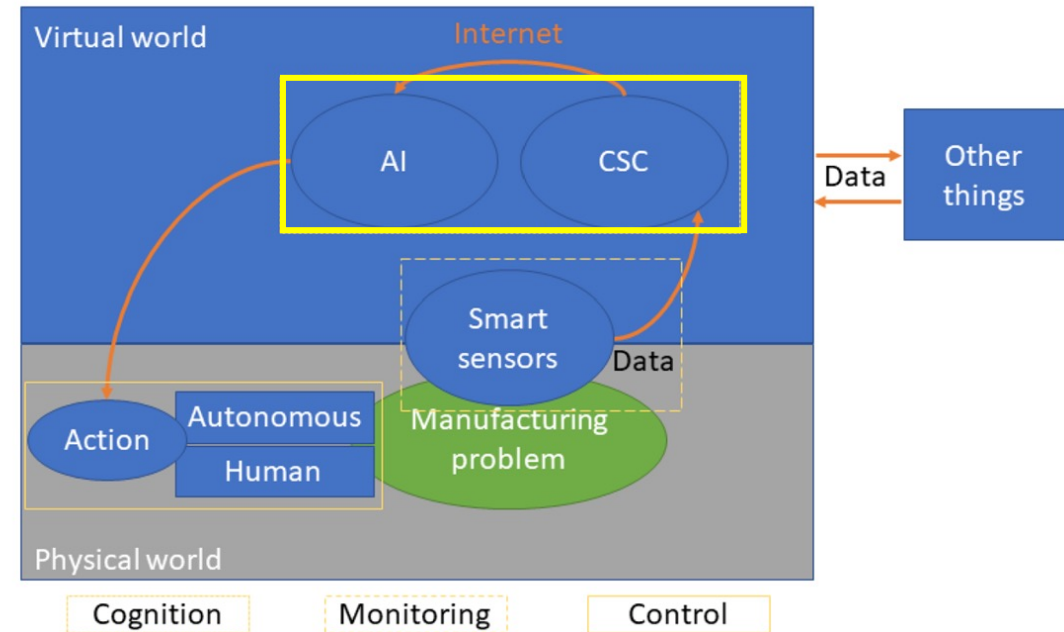
- “things”, with embedded sensors are deployed to generate the capacity of observing the system
- Quality characteristics of interest of 100% of the manufacturing items are observed
- Characteristics may be at a process, product, material, or machine level



Industrial internet of things

Cognition

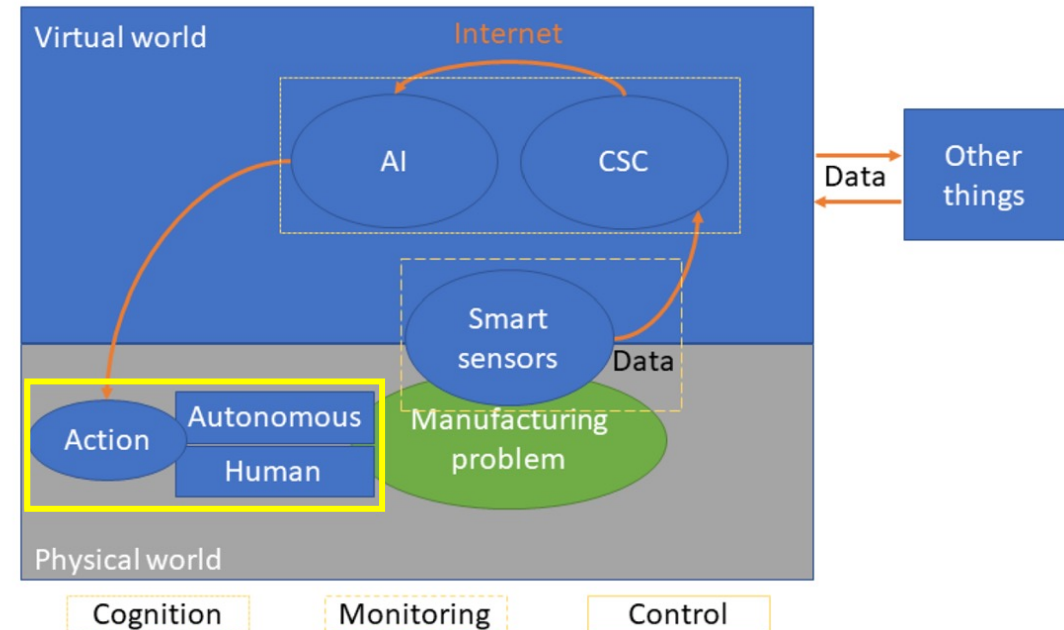
Using CSC technologies, AI is applied to the collected information learn the quality patterns and develop the classifier



Industrial internet of things

Control

- Classifier predicts, in real-time, the quality status of the manufacturing items
- Then, actions are taken to remove (rework or scrap) the defective items
- Autonomous actions rely on actuators, however, handling the detected defective items highly rely on human involvement



Chapter 3: The Data

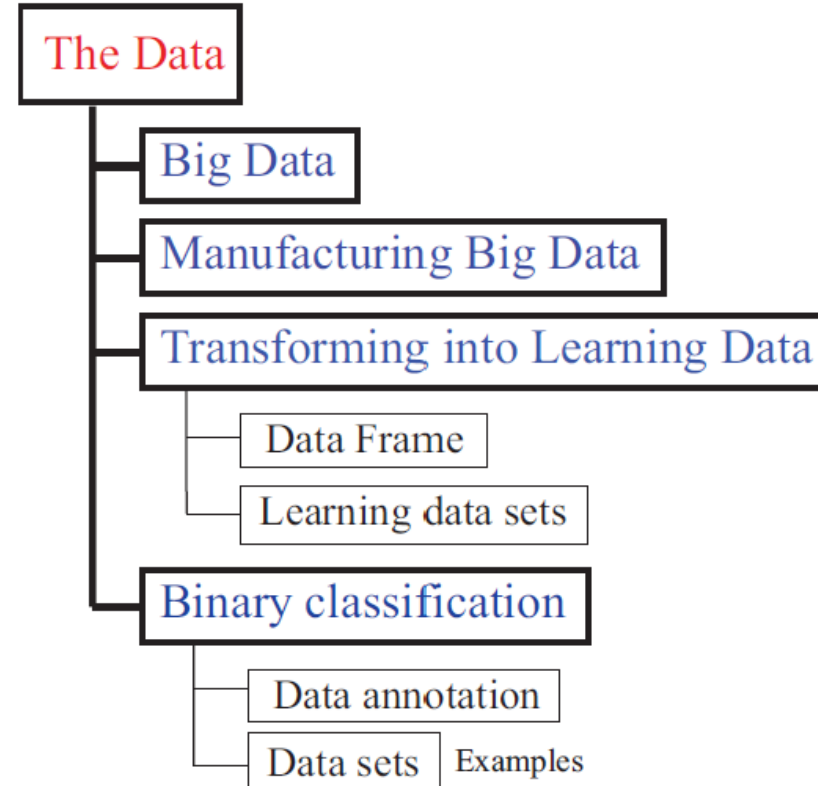
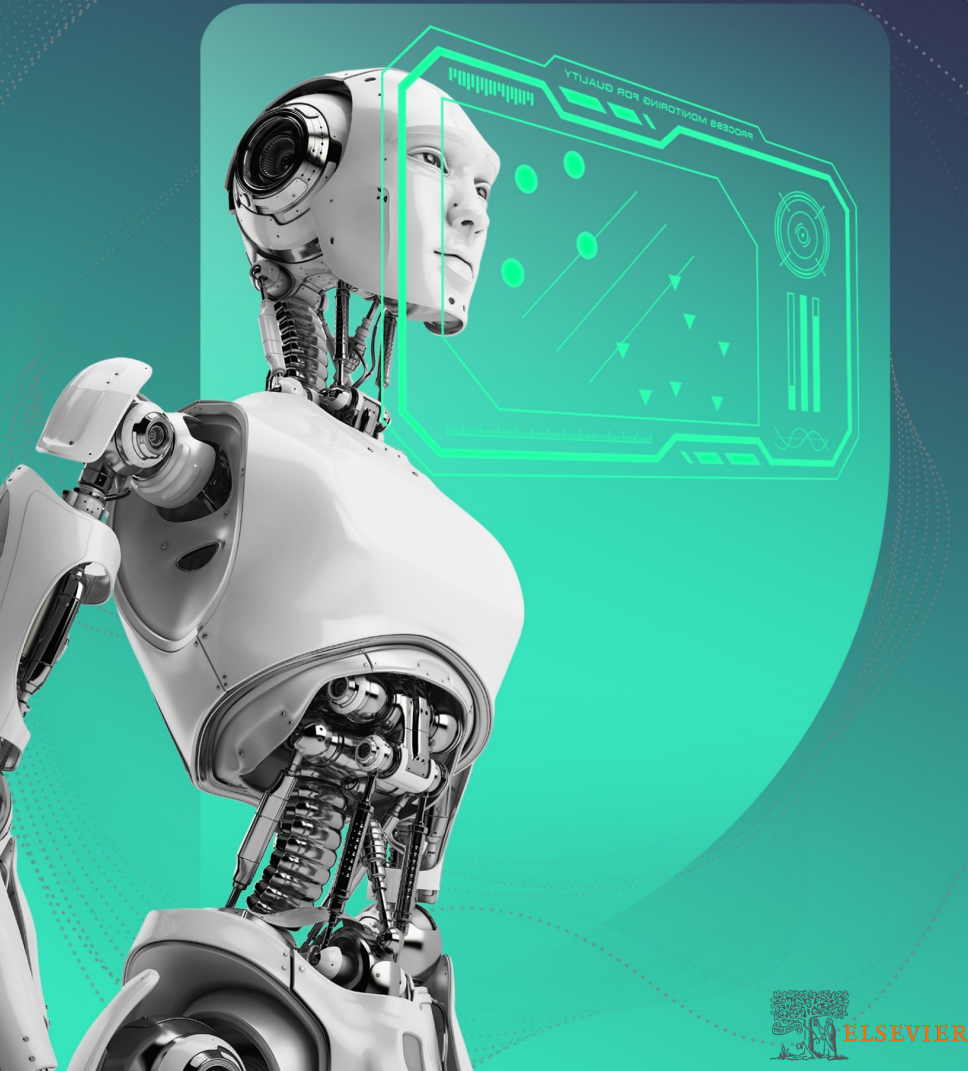
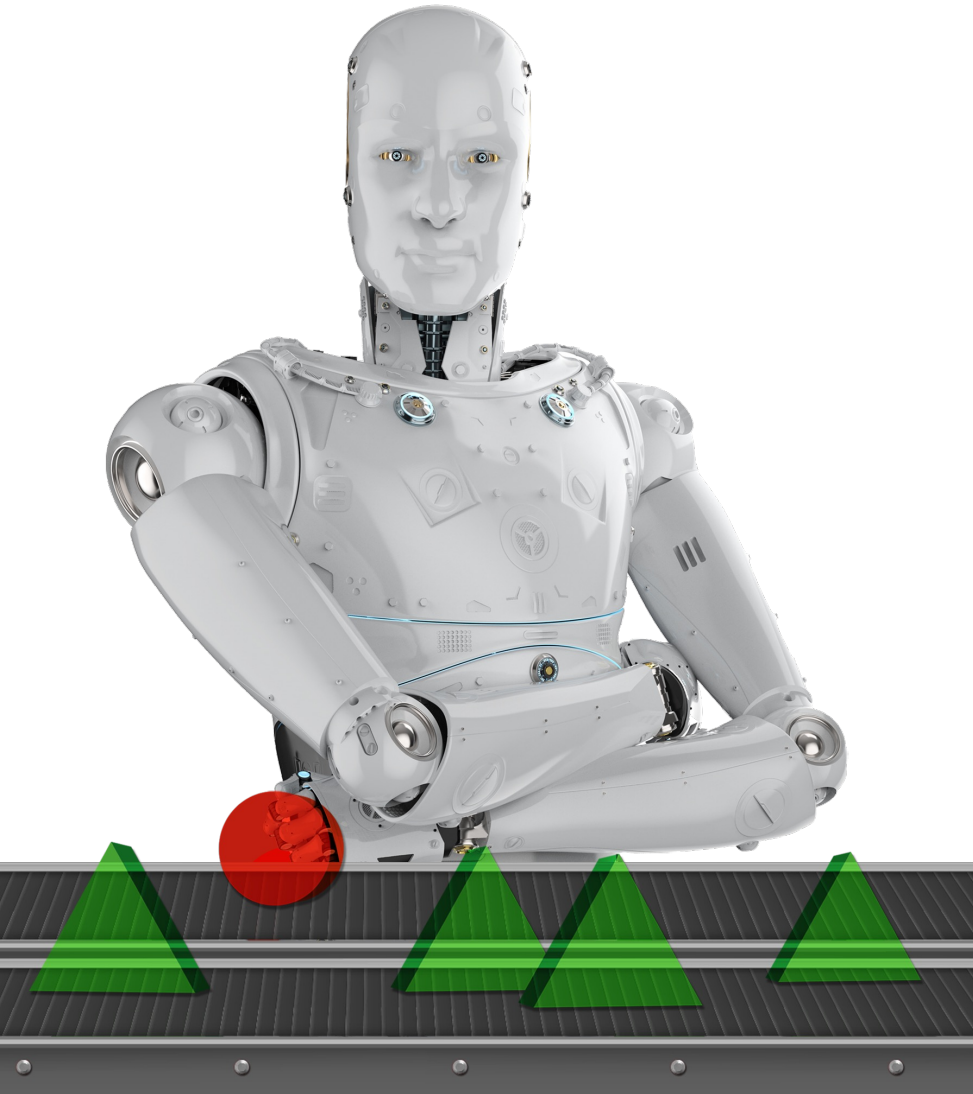


FIGURE 3.1 Contents of the Chapter 3



ELSEVIER



Big data

Big data

- Big data refers to data sets that are too large or complex to be studied and manipulated using traditional data analysis software such as Microsoft Excel following
- RAM of personal computer might be too small to handle big data
- Even some statistical techniques are redefined to cope with the challenges posed by the data set sizes
- 79% of businesses agree that companies that don't embrace big data may lose their competitive advantage and might even go extinct



Big data

Definitions from TechTarget, Oracle, and SAS (respectively):

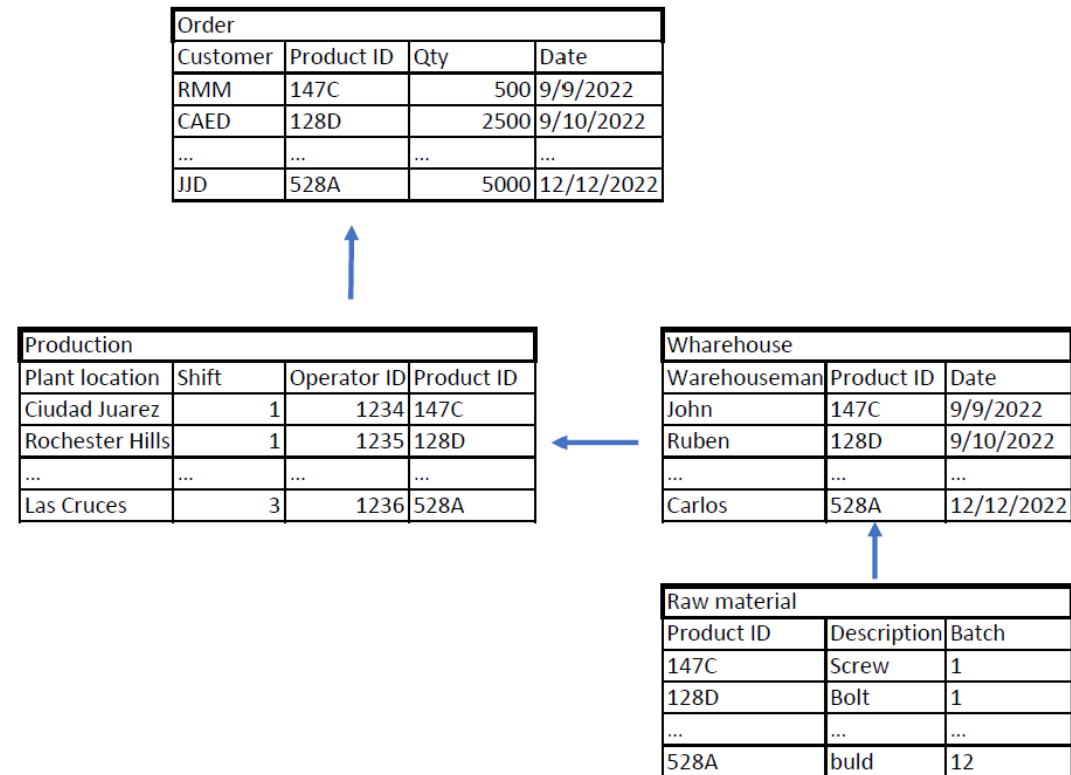
- ...a combination of structured and unstructured data collected by organizations that can be mined for information and used in machine learning projects, predictive modeling and other advanced analytics applications
- ...data that contains greater variety, arriving in increasing volumes and with more velocity
- ...a term that describes large, hard-to-manage volumes of data - both structured and unstructured - that inundate businesses on a day-to-day basis

Big data



Structured data

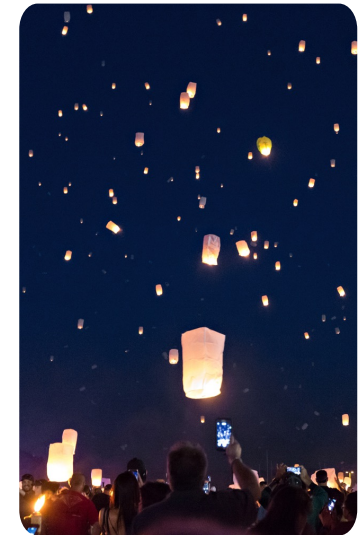
- Highly organized and formatted
- Includes patterns or unique identifiers that make it easily searchable in relational databases
- Database administrators can quickly input, search, and manipulate structured data by using a relational structured query language (SQL) database
- Stored in data warehouses



Big data

Unstructured data

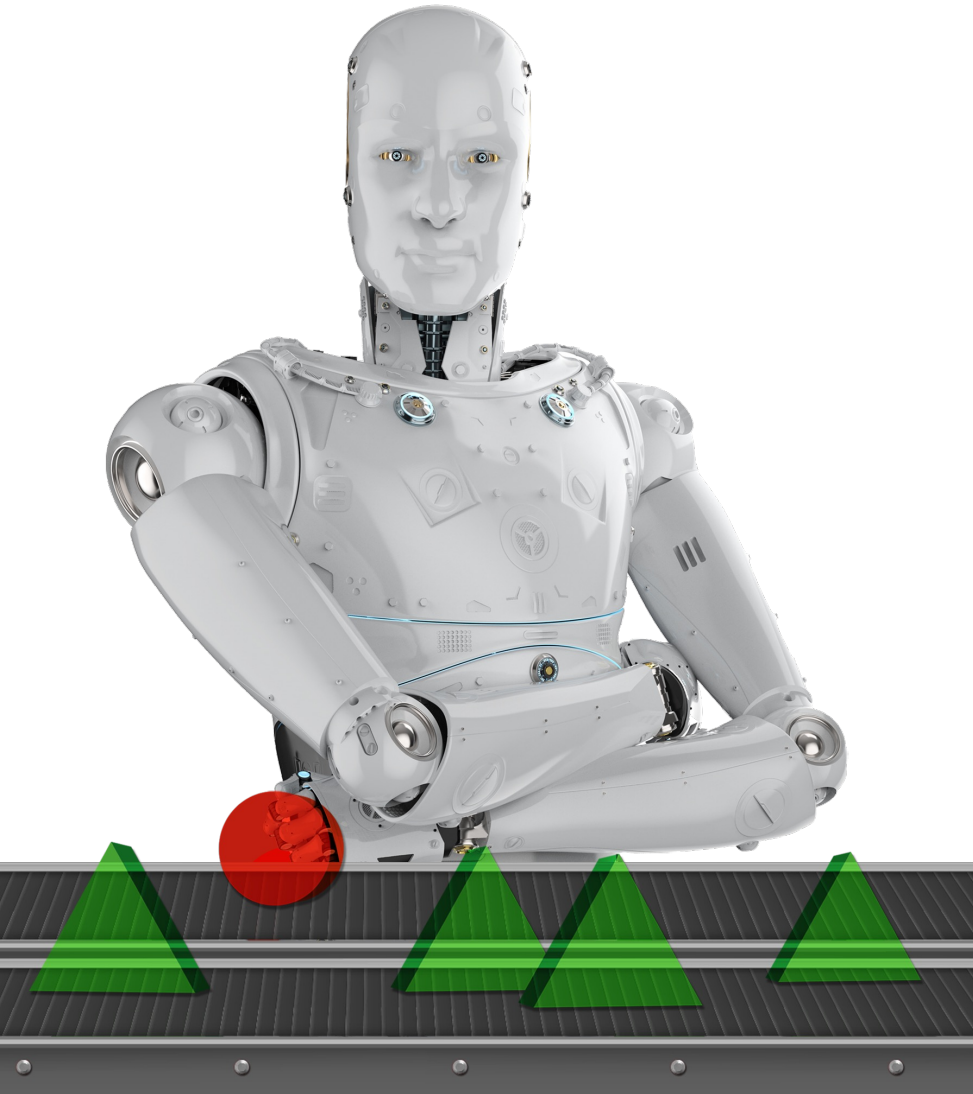
- No predefined format or organization
- More difficult to collect, process, and analyze
 - Audio
 - Video
 - Social media postings
- Stored in data lakes



Big data

Industrial big data

- Today, big data is being generated in all industries of the economy; science, media, sales, finance, healthcare, consumer products, marketing, and service
- In 2012, GE created a subcategory for big data, industrial big data (IBD)
- Large amount of diversified time-series data generated by industrial machinery



Manufacturing big data (MBD)

Manufacturing big data

- Manufacturing big data (*MBD*) enables the intelligence of the *SM* systems
- Large amount of diversified time series generated by manufacturing systems and industrial equipment
- MBD is observational in nature
- Before presenting a data set to an algorithm for it to learn the patterns, each feature must be evaluated from an engineering or physics perspective to ensure no spurious patterns are learned

Manufacturing big data

Statistics

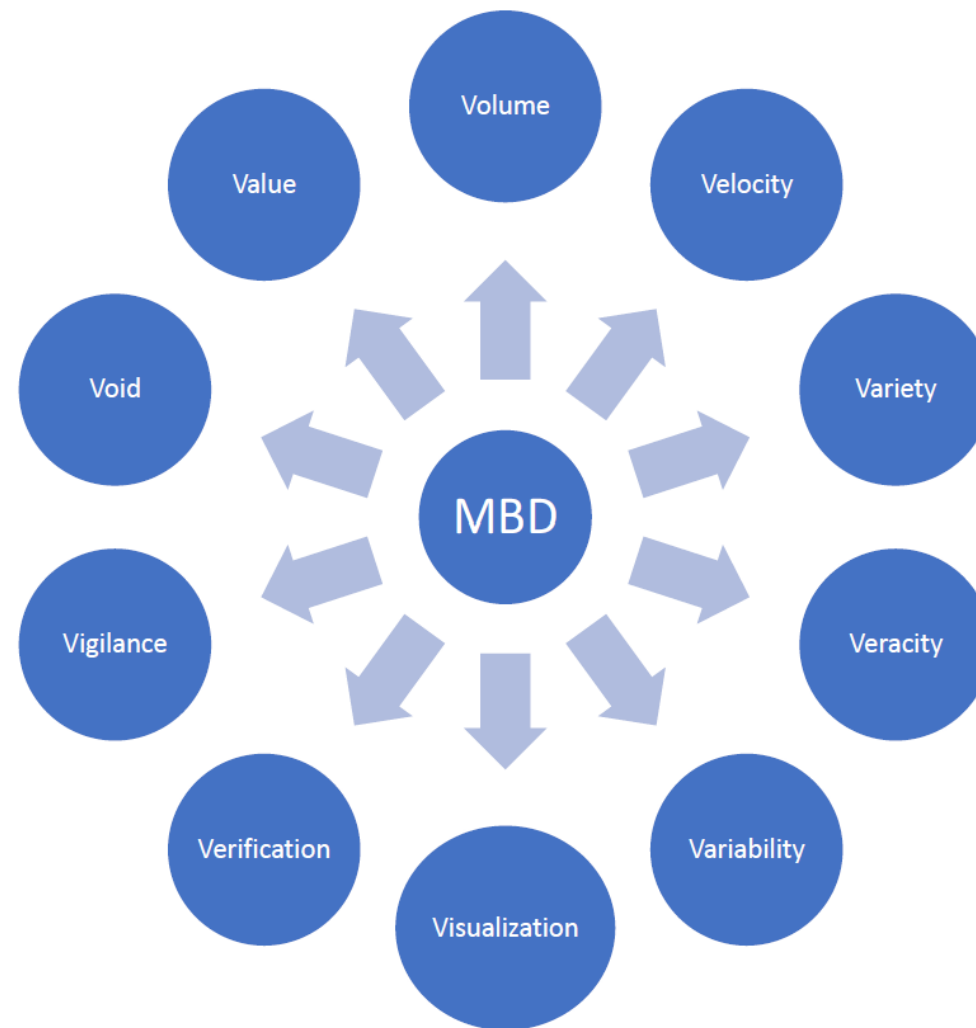
- Manufacturing industry generates more data than any other sector of the economy
- Only few companies are harnessing it
- One oil-and-gas company, discards 99 percent of its data before decision makers have a chance to use it
- Global managers and leaders in both business and IT reported that around 60 % of their data remains dark; unanalyzed, unorganized, and unused
- Another study found that up to 73 % of company data goes unused

Manufacturing big data

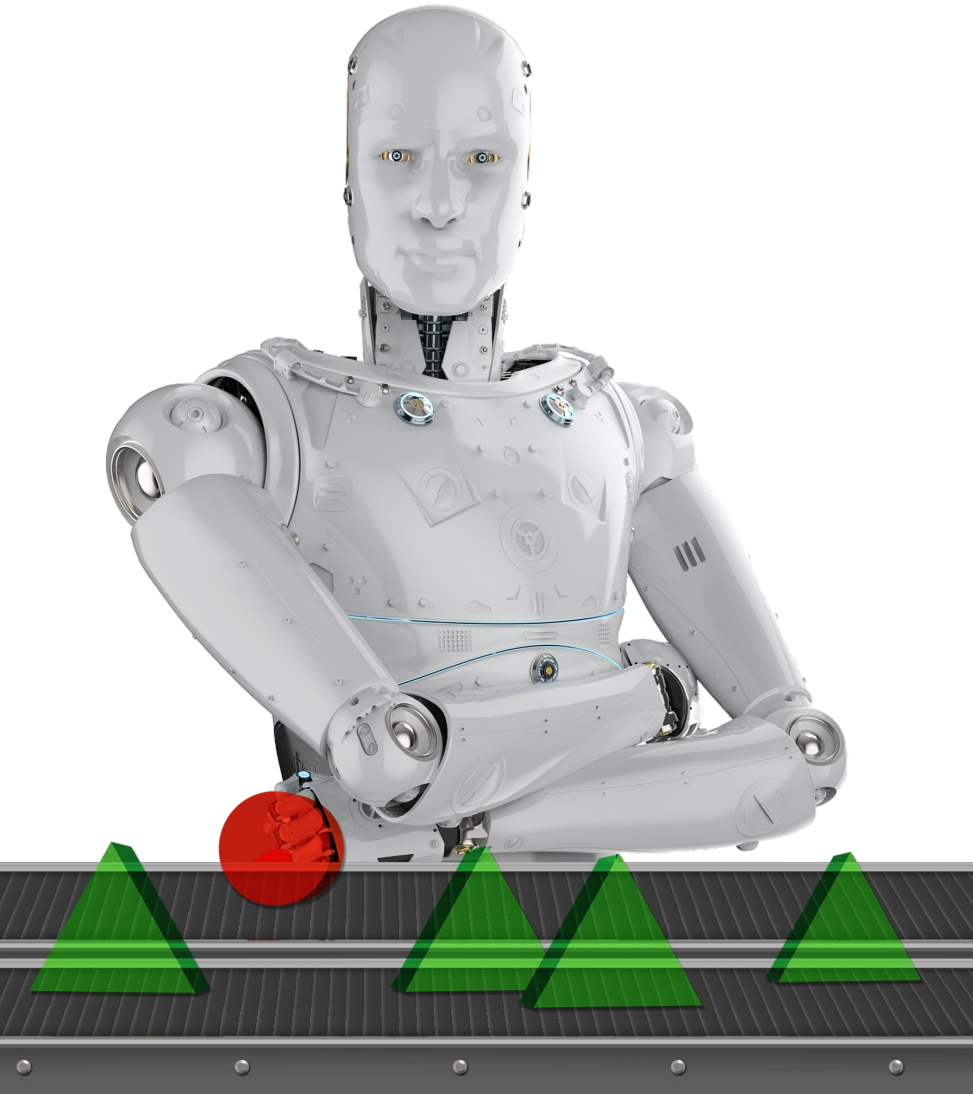
Important consideration

- Although today is relatively easy and cheap to generate and store data, it is important to define the project and the required data to solve it, before indiscriminately creating databases
- Storing large volumes of data with low levels of relevant information is not a good idea
- Maintaining the ever-increasing volumes of data will start generating monthly cloud storing bills
- Adding ever-increasing costs to the Q 4.0 initiative
- There is no need for a large up-front investment in databases without a clear problem definition
- Instead, the business value and the expected intelligence the data will generate should be defined in advance

10 V's



Manufacturing
big data



Transforming big data into learning data

Transforming big data into learning data

Data frame

- Structured data is stored in the form of tables
- Tables are merged in a data frame using languages like Python and R
- Data frame is a two-dimensional labeled data structure aligned in a tabular fashion (rows, columns)
- Each column contains values of one variable and each row (sample) contains one set of values from each column
- Data frame characteristics
 - Column names should be non-empty
 - Row names should be unique
 - Data can be of numeric, factor, or character type
 - Each column should contain same number of data items

Transforming big data into learning data

- Data frames may include non-predictive variables, such as indexes or the unique identifiers, used to merge the data from different tables
- Important to study the information contained in each column to derive the learning data set
 - Eliminate non-predictive variables
 - Avoid data leakage

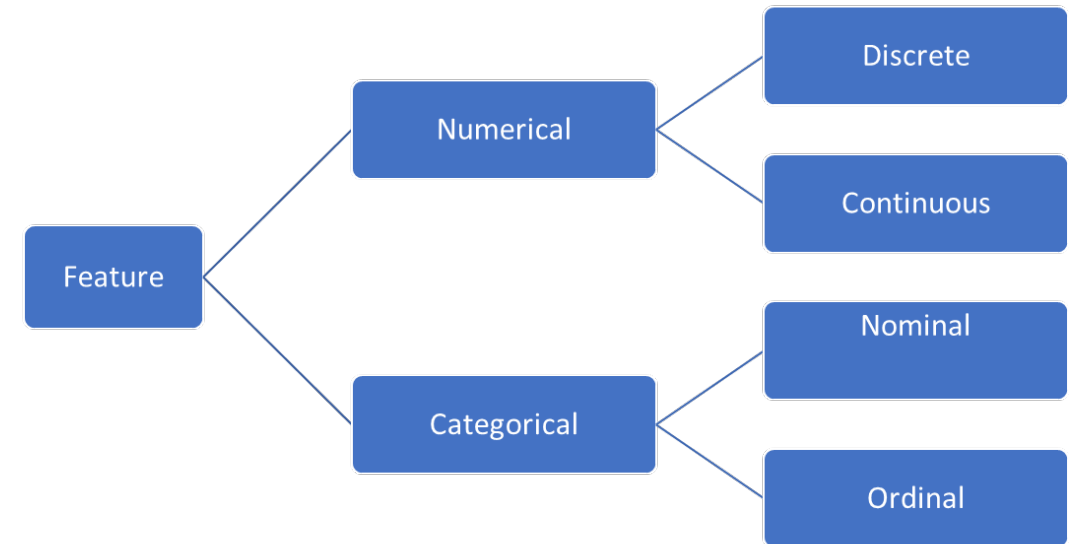
unique_ID	Feature space									Label vector	
	feature_1	feature_2	feature_3	feature_4	feature_5	feature_6	feature_7	feature_8	feature_9	label	
29511	-7.680000e-06	-1.780000e-05	-0.000261	0.000003	7.020000e-05	0.000130	0.014212	0.014230	0.014491	1	Samples
24952	-1.190000e-05	-5.650000e-05	0.000294	-0.000002	2.570000e-06	-0.000201	0.023739	0.023796	0.023501	0	
42380	-2.260000e-05	-8.100000e-05	-0.000164	-0.000012	-3.970000e-05	-0.000082	0.013083	0.013164	0.013327	0	
27008	8.940000e-07	-1.410000e-07	-0.000007	-0.000002	-1.820000e-07	0.000006	0.012224	0.012224	0.012231	0	
15661	2.600000e-06	3.100000e-05	0.000231	0.000006	5.740000e-05	-0.000193	0.004002	0.003971	0.003740	1	

Non-predictive variable

Transforming big data into learning data

Types data

- Structured learning data set is a preprocessed matrix derived from the data frame that includes feature vectors (i.e., predictive variables)
- Feature vector is an observable variable of the phenomenon being investigated
- There are two main types of features, numerical and categorical



Transforming big data into learning data

- Numerical, these type of features are usually direct process measurements with a physical meaning
- Categorical, this type of features represents process characteristics or categories
- Ordinal features are a mix of numerical and categorical
- Most MLAs can only learn from numerical values
- Important to determine if is necessary to transform the categorical into numerical features before applying the MLA
- Transformation process is called feature encoding

Sample	Pressure (psi)	Temp (fahrenheit)
1	30.25	2551
2	31.85	2550
3	33.54	2700
4	35.00	2596
5	34.98	2598
6	34.5	2605
7	33.01	2698
8	31.01	2601
9	30.56	2595
10	32.67	2588

Numerical

Machine	Machine2	Machine3
3	0	1
2	1	0
3	0	1
2	1	0
1	0	0

Categorical

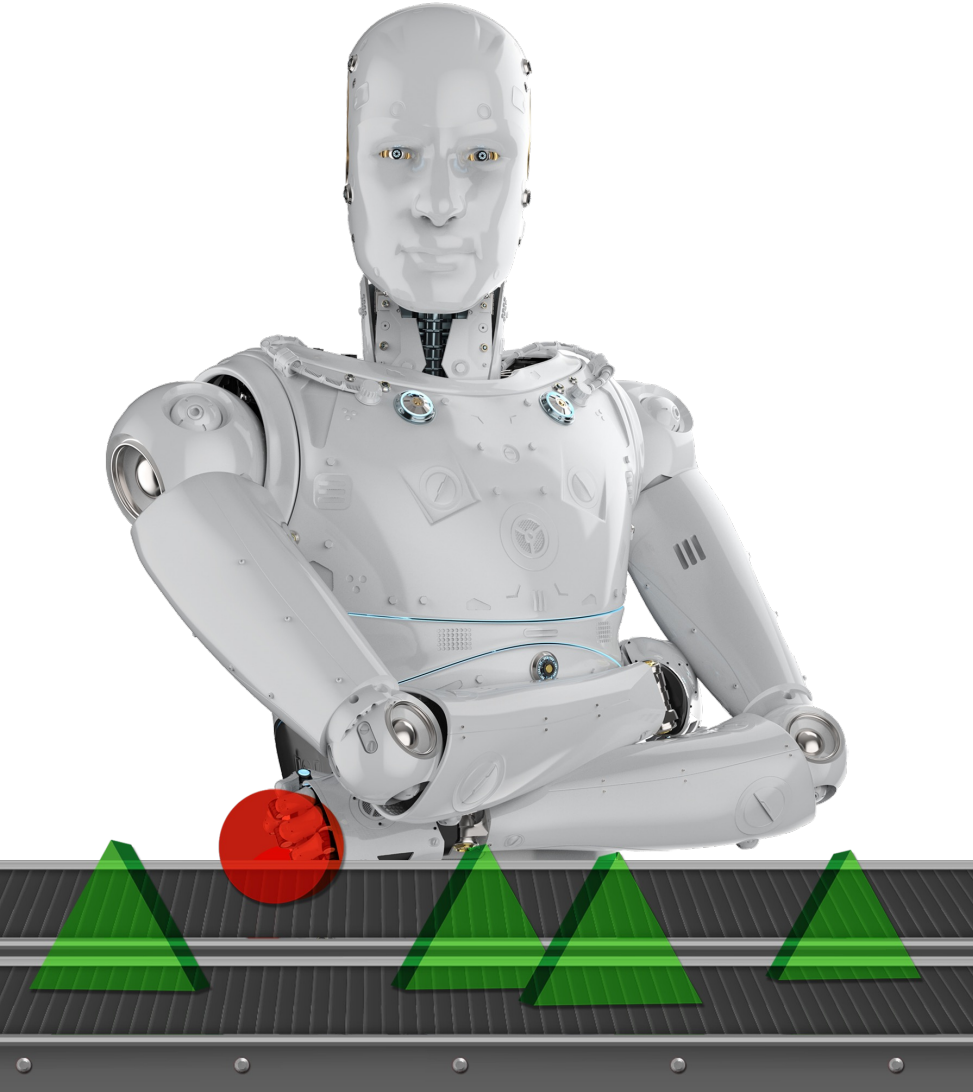
Treatment	Treatment
Yes	1
No	0
Yes	1
No	0

Categorical - binary

Size	Size
large	3
small	1
medium	2
large	3

Ordinal

Binary classification of quality (BCoQ) data sets



Binary classification of quality data sets

- Binary Classification of Quality (BCoQ) data sets, refer to the process-derived data generated and collected with the objective of applying ML or DL algorithms to develop a binary classifier
- Each sample has two potential outcomes:

$$QualityLabel = \begin{cases} 1 & \text{if } i^{th} \text{ item is } \textit{defective} (+) \\ 0 & \text{if } i^{th} \text{ item is } \textit{good} (-) \end{cases} \quad \left. \begin{array}{c} \text{red circle} \\ \text{green triangle} \end{array} \right\} \text{Color convention}$$

Notation

- Positive label or result refers to a defective item, a negative label or result refers to a good quality item

Binary classification of quality data sets

Characteristics

- Observed outcome of each sample is binary label (i.e., good, defective)
- Highly/ultra imbalanced (minority/defective class count $< 1\%$)
- Hyper-dimensional features spaces, including relevant, irrelevant, trivial, and redundant features
- Mix of numerical and categorical variables (i.e., nominal, ordinal or dichotomous)
- Different engineering scales
- Incomplete data sets
- Time dependent
- Transient sources of variation

Binary classification of quality data sets

Data annotation

- Labeling the quality class of each sample is usually a **manual annotation** process
- **Engineering evidence** is used and compared with the specifications to determine if the quality is met
- Labeling requires **high accuracy**, since the classifier will use this logic (or learned patterns) to automatically predict quality events in production

Binary classification of quality data sets

BCoQ data sets examples

Unstructured – signals

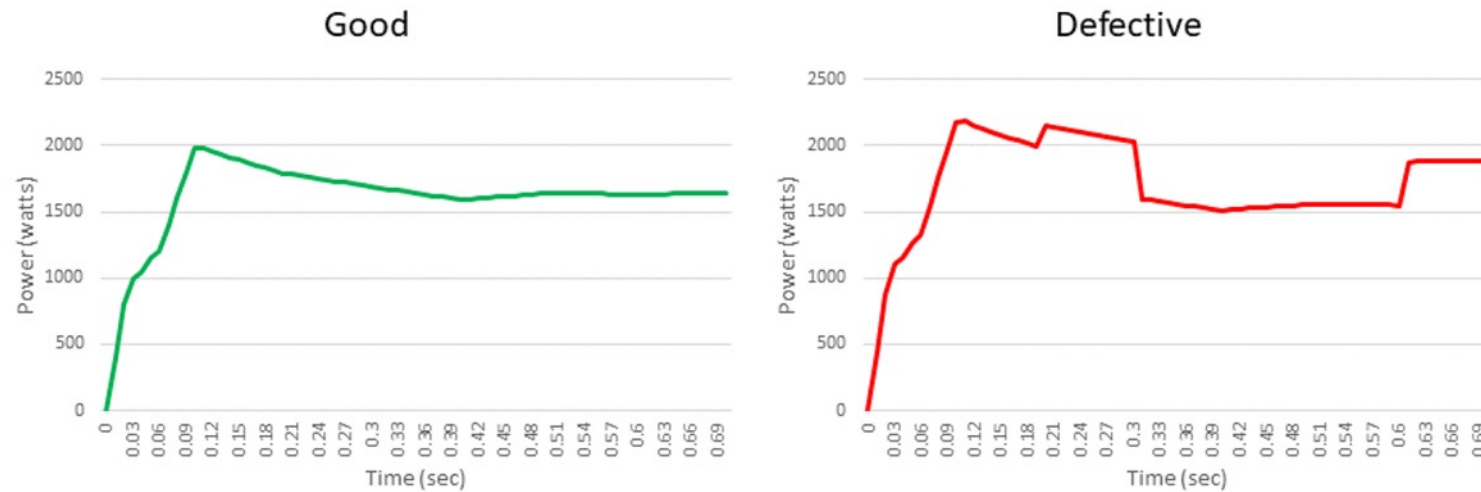


FIGURE 3.5 Signal from a power sensor (low resolution).

Binary classification of

Unstructured – images

- Vision systems for quality control are one of the most important research topics in Quality 4.0
- Enable the elimination or reduction of visual inspections through the application of DL techniques



FIGURE 3.6 A learning data set aimed at detecting **misalignments** (e.g., in the car body).



FIGURE 3.7 A learning data set aimed at detecting a **missing component** (e.g., tire).



FIGURE 3.8 A learning data set aimed at detecting a **common defect**.

Binary classification of quality data sets

Structured

Sample	Signal derived			Label
	mean(signal)	std(signal)	max(signal)	
1	Derived value	Derived value	Derived value	0
2	Derived value	Derived value	Derived value	0
3	Derived value	Derived value	Derived value	0
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
m	Derived value	Derived value	Derived value	1

FIGURE 3.9 A learning data set based on features derived from a signal.

Sample	Process features			Label
	Pressure	Speed	Humidity	
1	Observed value	Observed value	Observed value	0
2	Observed value	Observed value	Observed value	0
3	Observed value	Observed value	Observed value	1
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
m	Observed value	Observed value	Observed value	0

FIGURE 3.10 A learning data set based on process measurements.

Binary classification of quality data sets

Structured

Sample	Coordinates			Label
	x	y	z	
1	Observed value	Observed value	Observed value	0
2	Observed value	Observed value	Observed value	1
3	Observed value	Observed value	Observed value	0
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
m	Observed value	Observed value	Observed value	0

FIGURE 3.11 A learning data set based on deviations from nominal.

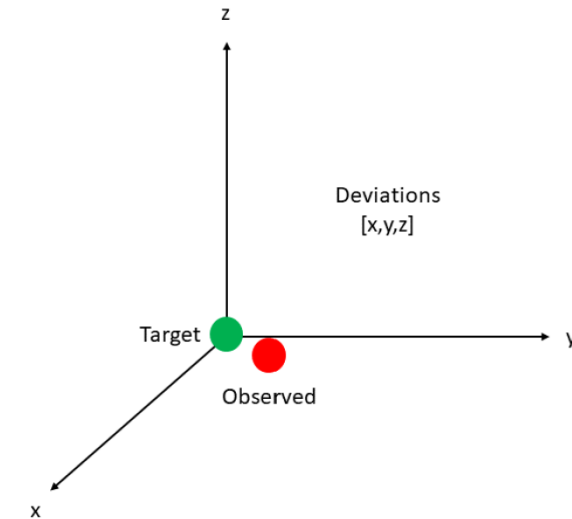


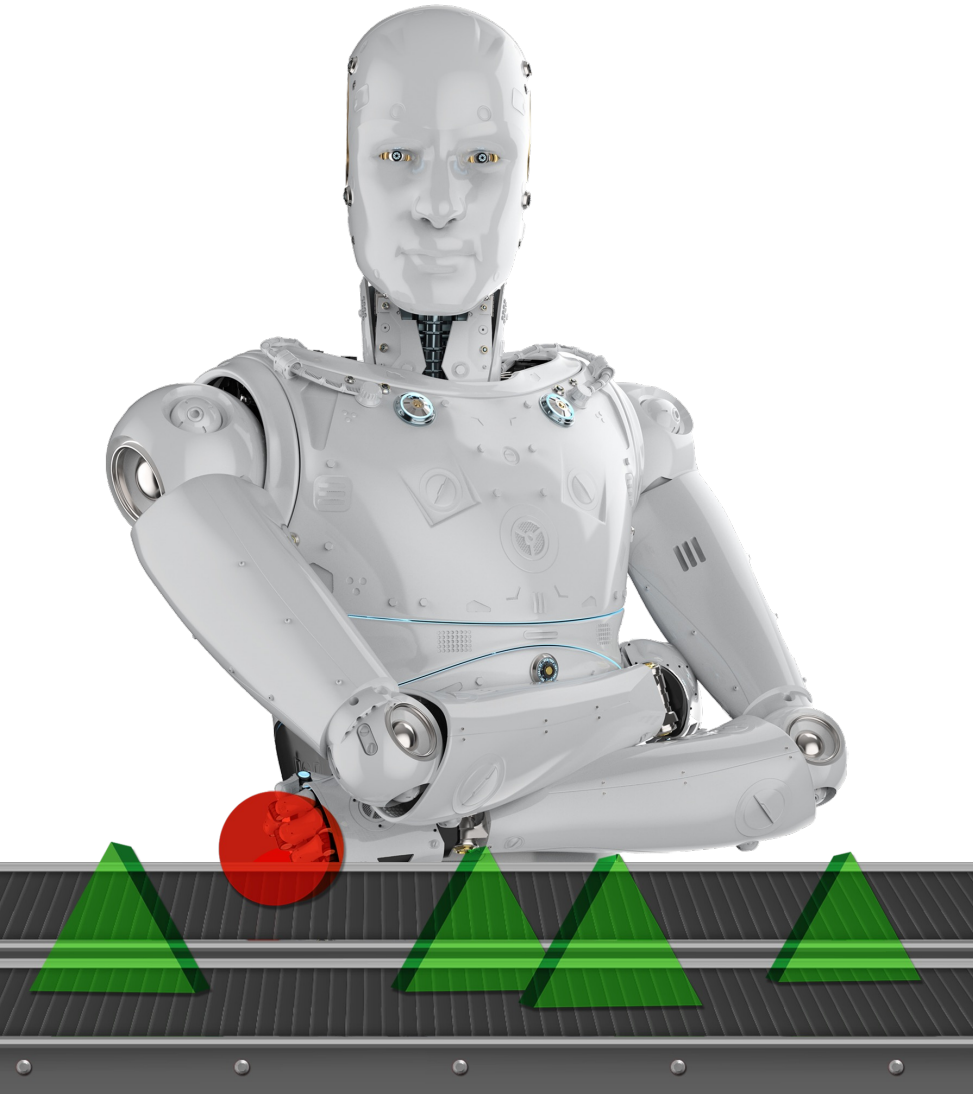
FIGURE 3.12 Target and observed position. Deviations in a 3-dimensional space (x,y,z).

Quick discussion

Think of a quality problem, which data can you potentially used to solve it?

Is it structured or unstructured?

What is the type of data?



Introduction to project selection

Activity 3

Access the following links and derive 2-3 remarks, 10mins

<https://towardsdatascience.com/how-to-say-no-to-projects-d6f88641ab3c>

<https://towardsdatascience.com/how-to-choose-which-data-projects-to-work-on-c6b8310ac04e>

<https://hbr.org/2018/10/how-to-decide-which-data-science-projects-to-pursue>

<https://hbr.org/2019/02/how-to-choose-your-first-ai-project>

Discussions, 5mins

Activity 4

Homework

Identify two to three potential projects.

- Write a report describing:
 - Process
 - Features
 - Problem
 - Value

The image features a classic 'The End' title card. The text 'The End' is written in a white, elegant, cursive script font, centered within a solid black circle. This central circle is surrounded by several concentric circles of varying shades of gray, creating a tunnel-like or ripple effect. The overall composition is symmetrical and visually striking.

The End

Activity 1

Read your assigned source and derive 2-3 remarks, 10mins

- <https://www.businesswire.com/news/home/20190102005073/en/NewVantage-Partners-Releases-2019-Big-Data-and-AI-Executive-Survey>
- <https://dlabs.ai/blog/top-10-challenges-companies-face-during-ai-adoption/>
- <https://www.inc.com/jeff-barrett/misusing-data-could-be-costing-your-business-heres-how.html>
- <https://www.forbes.com/sites/forbesbusinesscouncil/2022/11/21/the-top-five-ways-ai-is-transforming-business/?sh=6ff283378e7f>
- <https://www.forbes.com/sites/louiscolumbus/2019/03/17/machine-learning-engineer-is-the-best-job-in-the-u-s-according-to-indeed/?sh=267fec37bb0a>

Discussion 10mins

Appendix

<https://www.cpajournal.com/2021/08/10/can-robotic-process-automation-improve-quality-control-in-audits/>

<https://www.forbes.com/sites/forbestechcouncil/2023/01/06/beyond-innovation-overcoming-challenges-in-developing-and-deploying-ai-models/?sh=53a7496c56a5>

<https://www.automationworld.com/products/software/article/13315584/automation-opportunities-abound-for-quality-inspections>

<https://www.qualitymag.com/articles/96490-how-ai-software-enables-more-accurate-quality-inspection>

<https://www.forbes.com/sites/joemckendrick/2022/12/30/artificial-intelligence-without-the-right-data-is-just-artificial/?sh=22e3d860181b>

Breakdown of traditional quality control

Virtual case study

- 10 samples
- Three features
- Four defects
- Separate in three dimensions

Table 4. Virtual feature values and the associated quality status of each sample.

Index	Feature 1	Feature 2	Feature 3	Quality status
1	89	91	95	Defective
2	33	20	98	Good
3	95	93	88	Defective
4	40	45	56	Good
5	80	86	92	Defective
6	20	55	12	Good
7	52	60	80	Good
8	90	95	20	Good
9	99	89	60	Good
10	85	82	75	Defective

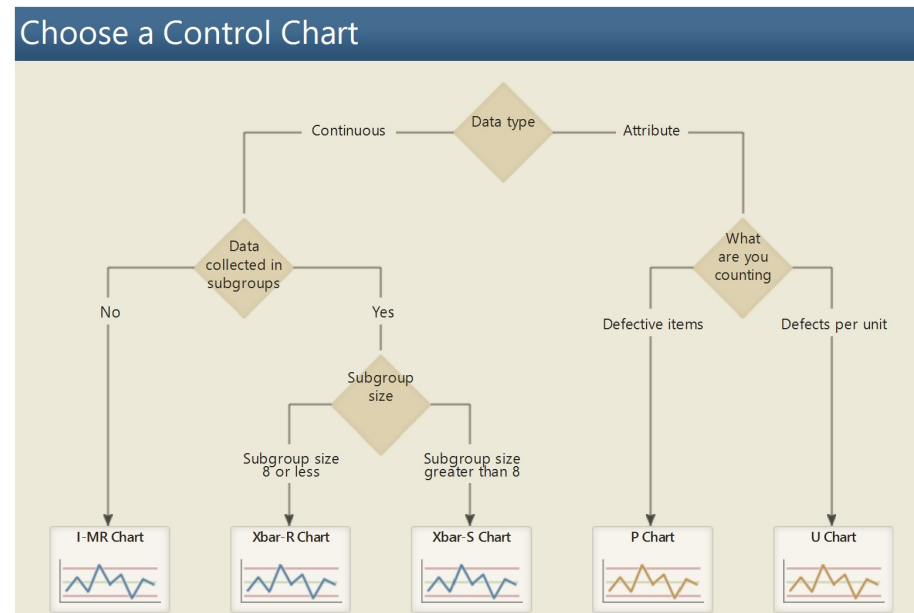


Figure 7. Decision tree for choosing a control chart (Minitab).



What is next?

Yellow Belt Certification



- Complete Course Comprehension
- Assigned Mentor Session
- Introduction to Project Selection

88% DISCOUNT
~~\$499.99~~ \$59.99

Buy Now

Green Belt Certification



3 Online Session Dates:

March 29
April 5
April 12

Time: 4:00 - 7:00 PM
Juarez time

- 3 LIVE Online Training Sessions
- Assigned Mentor
- Project Selection Advise

88% DISCOUNT
~~\$999.99~~ \$119.99

Buy Now



We also do on-site trainings

Breakdown of traditional quality control

- Univariate control charts
- No out of specifications detected

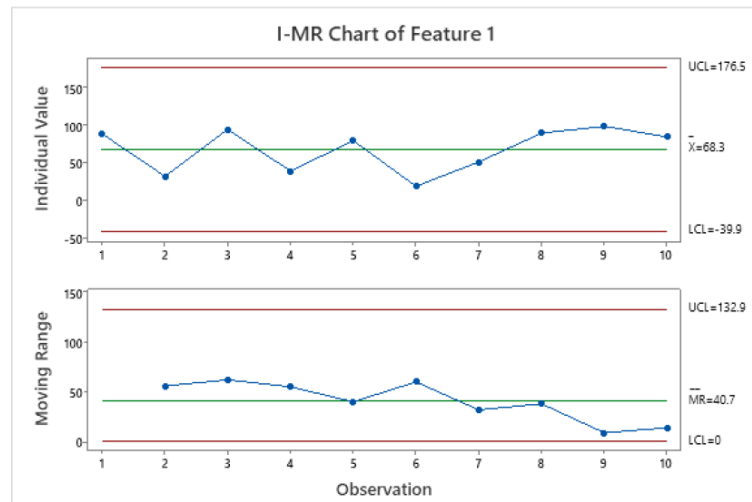


Figure 8. I-MR univariate control chart - Feature 1.

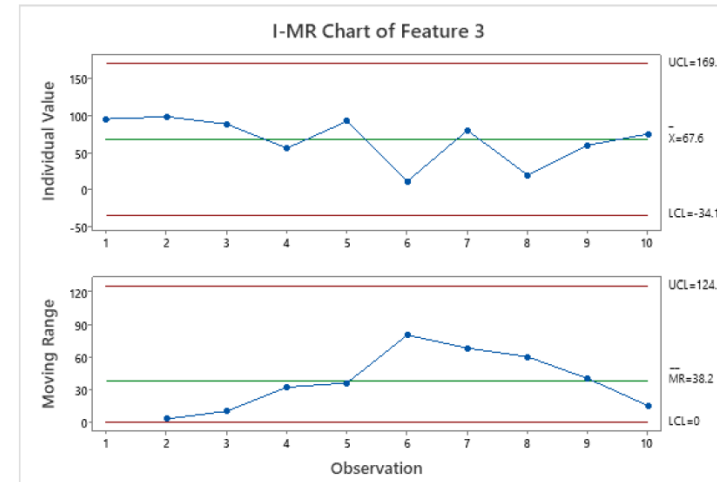


Figure 10. I-MR univariate control chart - Feature 3.

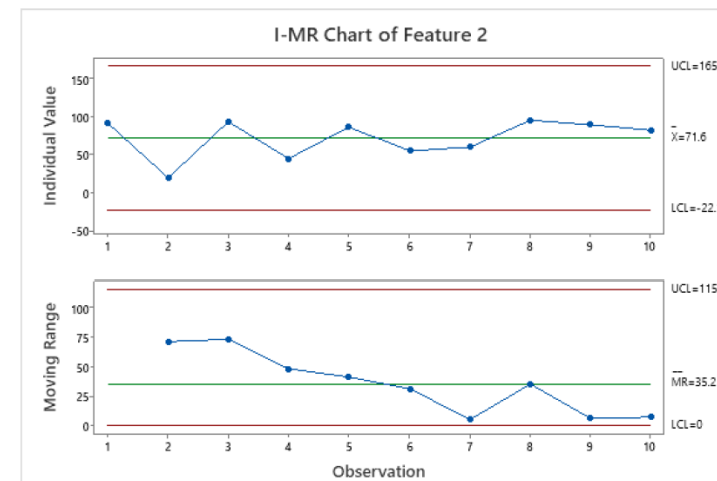


Figure 9. I-MR univariate control chart - Feature 2.

Breakdown of traditional quality control

- Multivariate control charts
- One FP each

Table 4. Virtual feature values and the associated quality status of each sample.

Index	Feature 1	Feature 2	Feature 3	Quality status
1	89	91	95	Defective
2	33	20	98	Good
3	95	93	88	Defective
4	40	45	56	Good
5	80	86	92	Defective
6	20	55	12	Good
7	52	60	80	Good
8	90	95	20	Good
9	99	89	60	Good
10	85	82	75	Defective

FP

FP

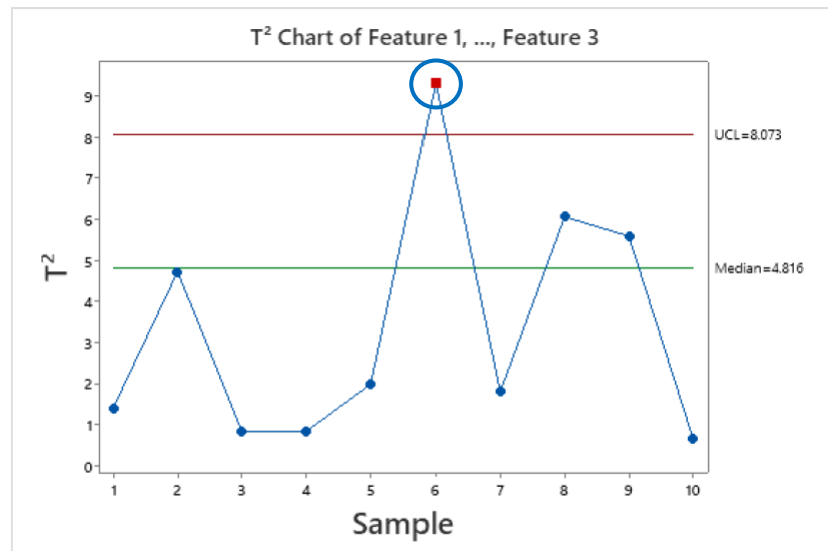


Figure 11. Multivariate Hotelling's T^2 - Feature 1-3.

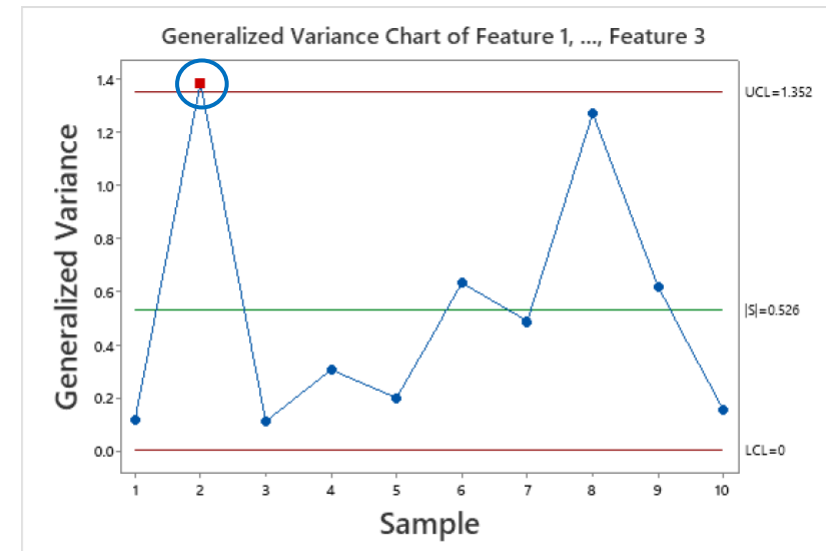
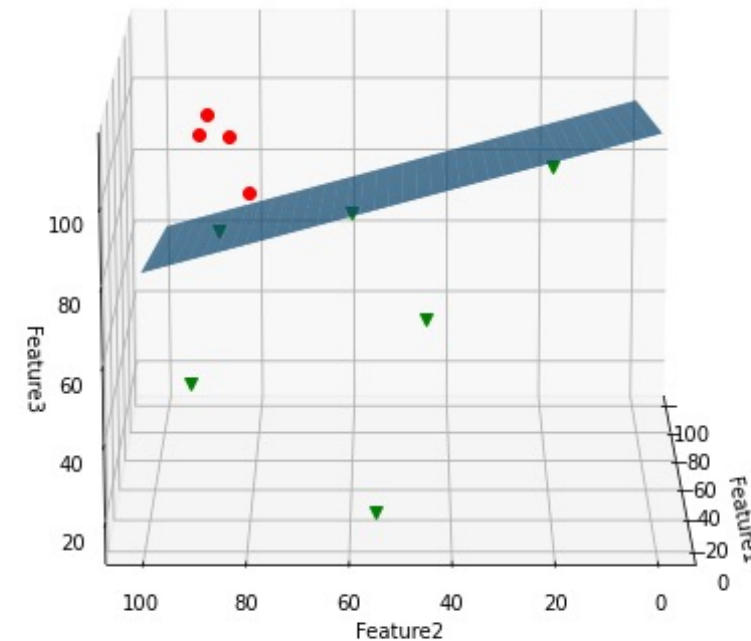
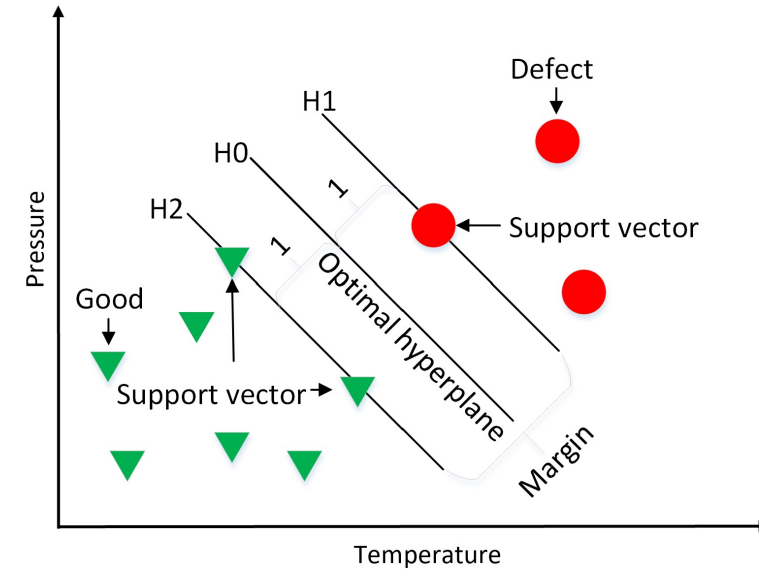
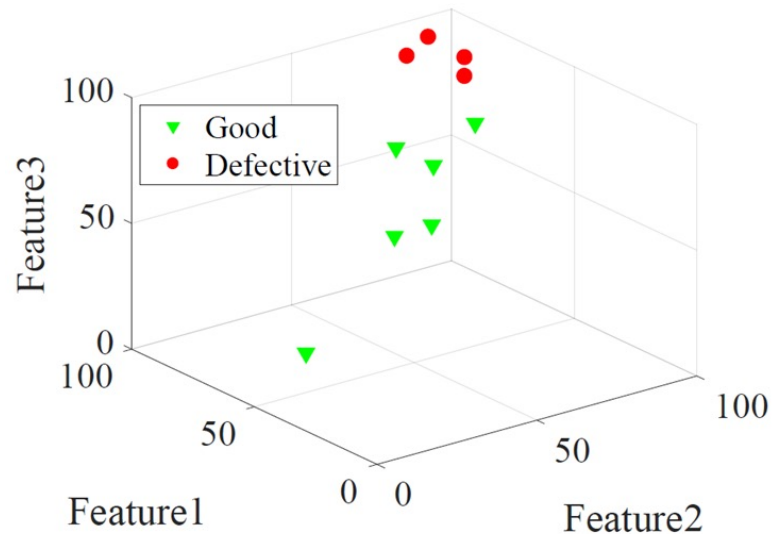


Figure 12. Multivariate generalized variance control charts - Feature 1-3.

Breakdown of traditional quality control

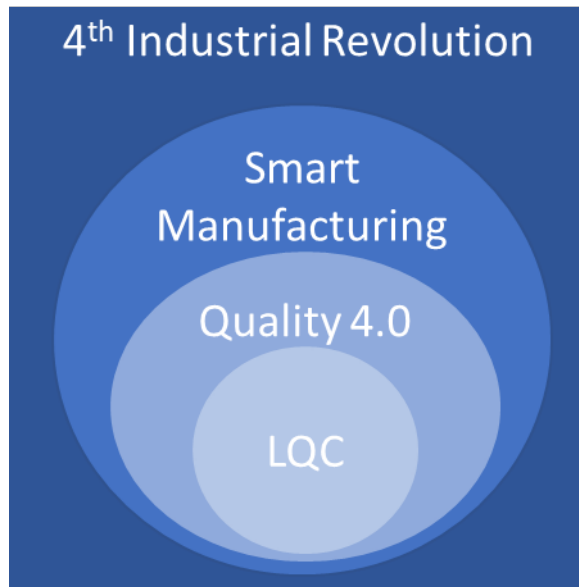
- Separating these classes is an easy task for the machine learning algorithms
- Support vector machine is a good option here



The rise of quality 4.0



LQC, a Quality 4.0 initiative



Evolution

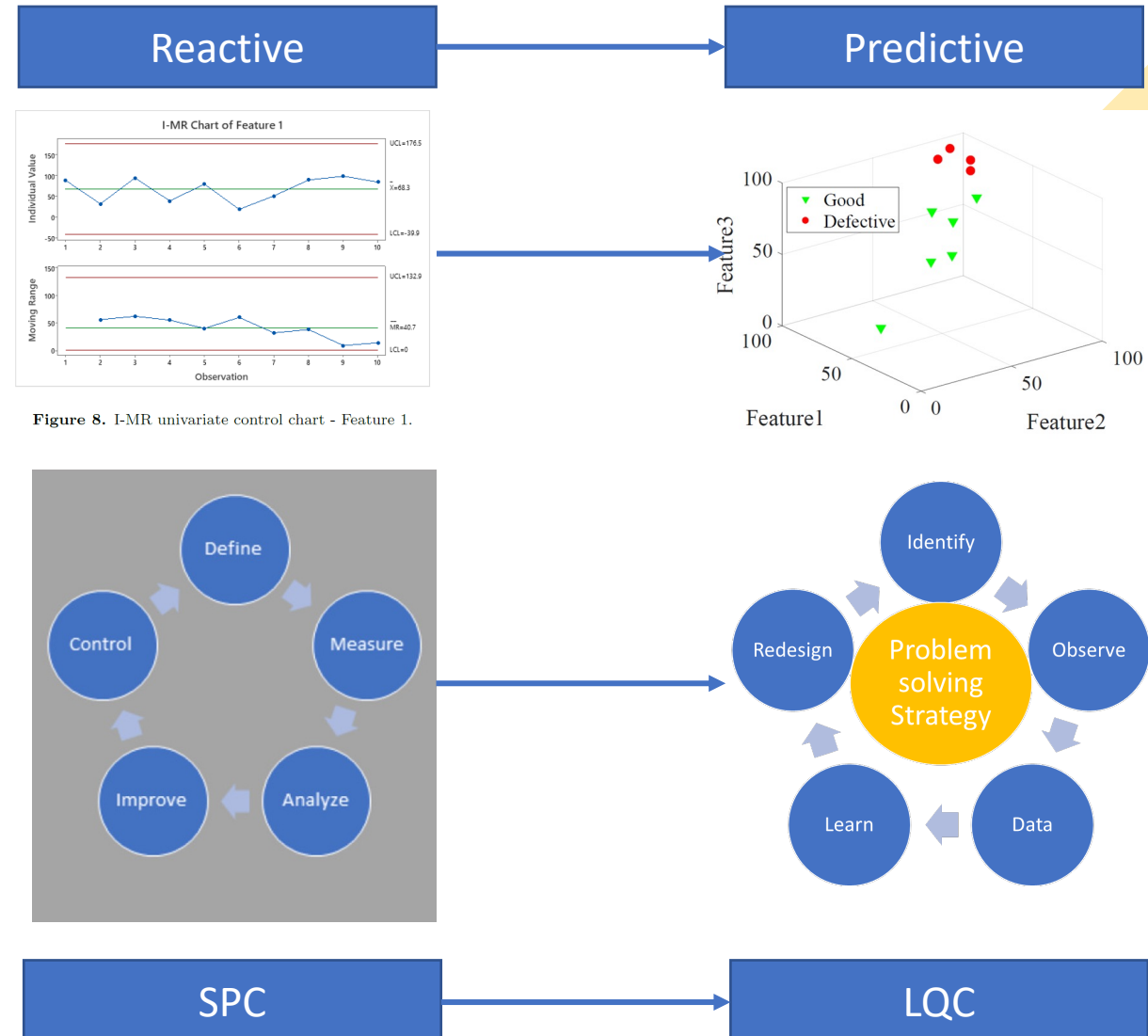


Figure 8. I-MR univariate control chart - Feature 1.



Quality 4.0

“The manufacturing companies to first implement a Quality 4.0 innovation system will be among the most advanced and influential companies in the world. Quality 4.0 combines the traditional methods and paradigms with the fourth industrial revolution technologies, such as AI, to create innovation in manufacturing. The first professionals to learn and deploy a Quality 4.0 initiative, will be leaders and advance rapidly to more important roles.”

Carlos A. Escobar, PhD

SHPE STAR of Today and Founder, Quality 4.0 Institute

Transforming big data into learning data

Learning data sets

- Integrates the predictive information with the labels of each sample
- Used to train, tune, and evaluate a ML model
- Recommended to split the learning data set following a time-ordered hold-out validation scheme.
- Split generates three different sub-sets



Transforming big data into learning data

- Training
 - Used to train the model
 - MLAs learn the patterns in this data set i.e., find the weights of the model
 - Largest data set
- Validation set
 - Used to change training process parameters (i.e., hyper-parameters) such as the number of neurons
 - Explore and evaluate different feature combinations
 - Stop the training process
 - Select the final model
 - Prediction ability, trust-ability from an engineering perspective, and parsimony are the factors considered when selecting the final model
- Test (hold-out)
 - Used to obtain unbiased prediction ability estimate of the final model
 - Neither used for training nor for tuning or selecting the final model
 - Plays an important role in ensuring the final model generalizes well to unseen data

Transforming big data into learning data

Recommendations on data splits

- Training data set must contain the patterns in the validation and test sets.

Hyper-parameters

- MLAs with small number of hyper-parameters are easier to validate and tune. Therefore, reduced size of validation set may be enough (e.g., 70-15-15)
- Larger validation set may be required for MLAs with many hyper-parameters (e.g., 70-20-10)
- Test set may not be required for MLAs with no hyper-parameters to tune (e.g., 70-30)



- If the data set is split into two sub-data sets, they are usually called training and test sets

Transforming big data into learning data

Data set size

- Relatively small data sets, in the order of hundreds or thousands, the validation and test set should be larger (e.g., 70-15-15)
- Validation and test sets may be reduced in huge data sets; in the order of hundred of thousands or millions (e.g., 90-5-5)
- Splitting the learning data into only two sets is not advisable when the MLA used to develop the model requires hyper-parameter tuning
- Final model may exhibit overoptimistic results

Binary classification of quality data sets

- Labeling process starts by defining an initial learning data set size
- Human annotation is performed
- Model is developed using the training and validation sets
- If the required accuracy is met:
 - Pilot runs are executed to evaluate the performance of the model in production
 - Otherwise, either more features, more data, or both is recommended

Amazon SageMaker

