SMART MANUFACTURING Forum

The Evolution of AI Applications for Process Control

Presenter: Jon Holt
Date: September 11, 2025

PDF/SOLUTIONS™

semi | SMART MANUFACTURING EXPO | TAIWAN

The Evolution of AI Applications for Process Control

SMART MANUFACTURING Forum

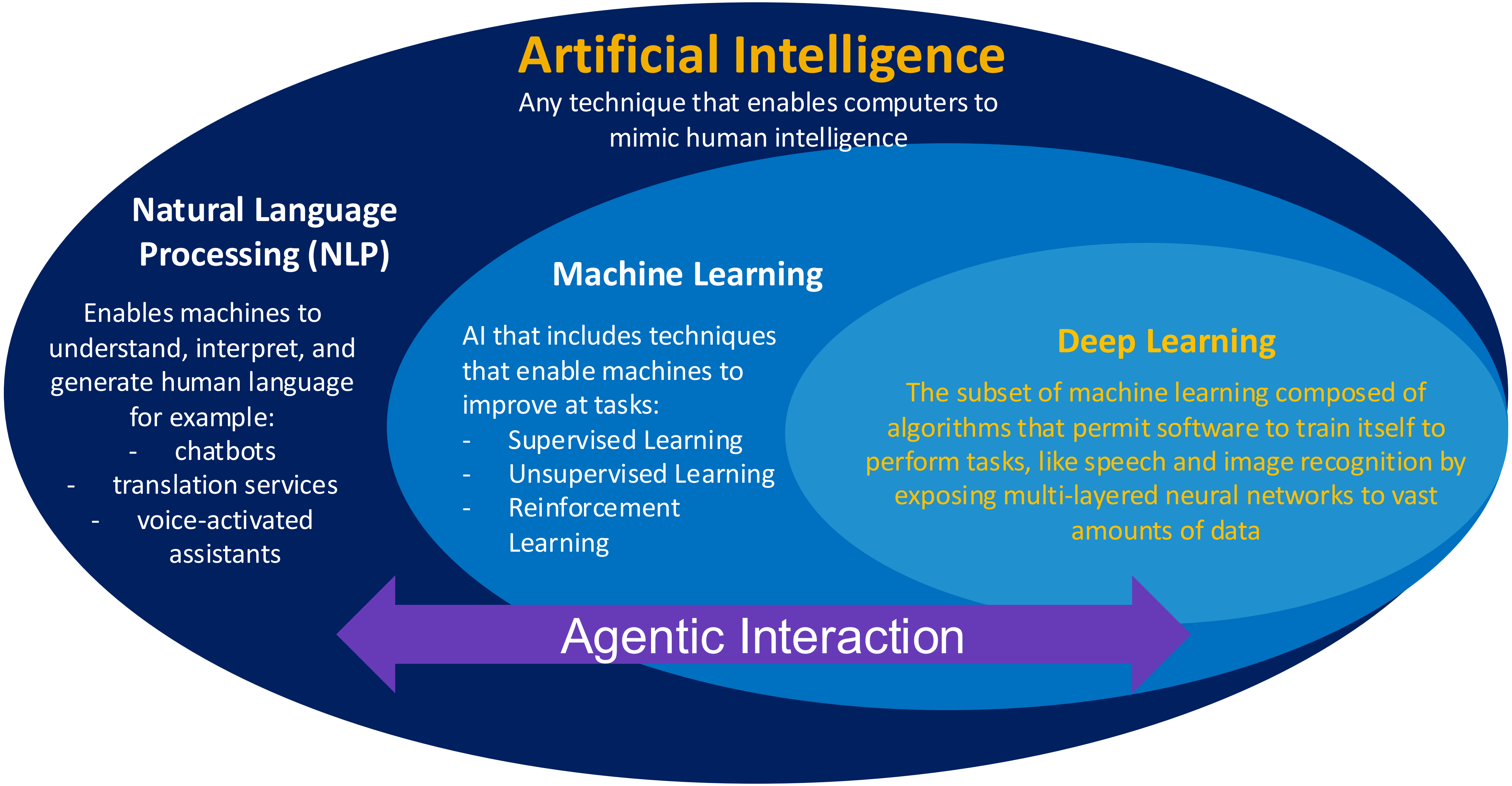
- **Artificial Intelligence** (AI) refers to the development of computer systems capable of performing tasks that typically require human intelligence, such as reasoning, learning, decision-making, and perception[1].

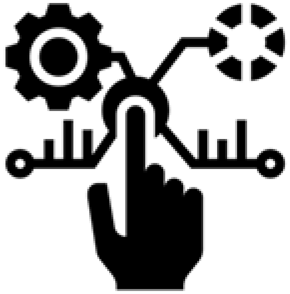
- Over the last **25+ years** AI applications have been successfully deployed for advanced process control in semiconductor manufacturing to enable variability reduction and improved operational cost and efficiency (SPC, APC, R2R, OCAPS, ML Models, etc.).

- Traditional applications, and new capabilities are being incorporated in Smart Manufacturing and Enterprise level control systems through the implementation of digital twins and novel AI capabilities to maximize ROI (Return on Investment) and enhance process control at all levels in manufacturing through **Agentic interaction**.

Artificial intelligence – Wikipedia
What Is Artificial Intelligence? Definition, Uses, and Types | Coursera

Machine Learning, Deep Learning, & Natural Language Processing





- *"AI can deliver significant business impact, but companies can maximize its value by taking an end-to-end approach. Weaving together strategy, process redesign, and human and technical capabilities, we create the fabric of an AI-driven organization, enabling the outcomes that drive businesses forward."*

AI @ Scale | AI Consulting and Strategy | BCG

Types of Artificial Intelligence

- **Reactive Machines:** These are the most basic type of AI systems that can only react to current scenarios without any memory of past events[3].
- **Limited Memory:** These systems can use past experiences to inform future decisions. Self-driving cars are an example[3].
- **Theory of Mind:** This type of AI, still in development, will understand emotions and human interactions[3].
- **Self-aware AI:** The most advanced form, which would possess self-awareness and consciousness. This remains theoretical[3].

Basic SPC

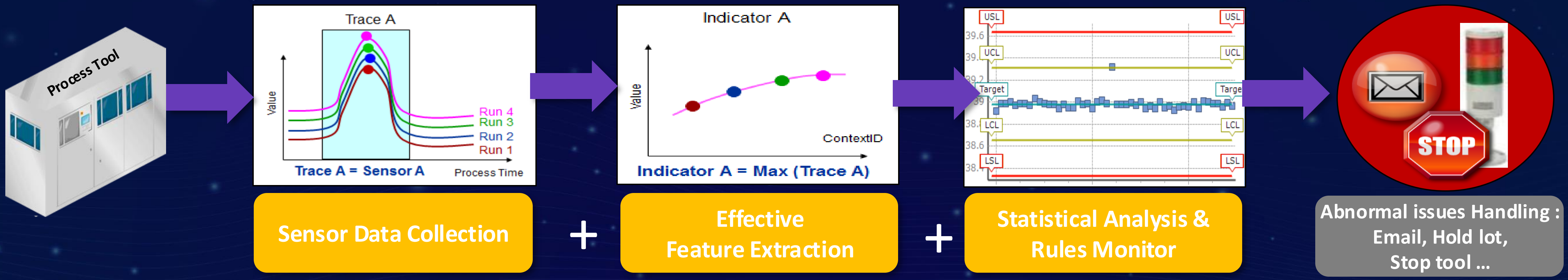
Most modern APC

In-Flight LLMs, GenAI

[3] What Is Artificial Intelligence? Definition, Uses, and Types | Coursera

Process Control AI Applications – Basic SPC

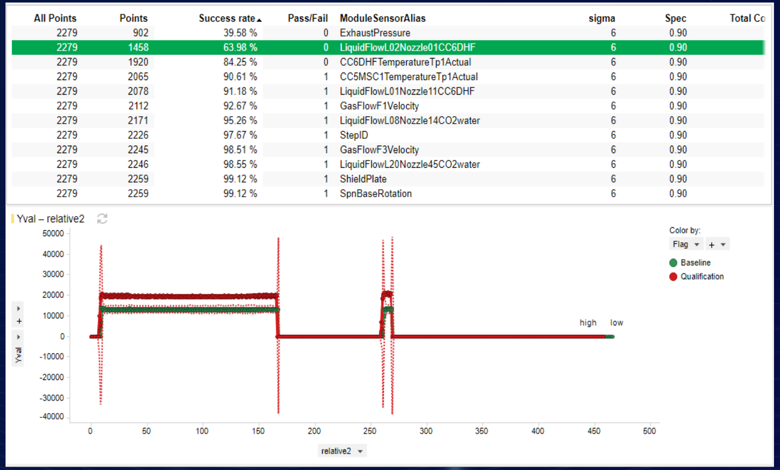
- Traditional Machine & Deep Learning: APC – Advanced Process Control, FDC – Fault Detection & Classification, SPC – Statistical Process Control, R2R – Run to Run control, VM – Virtual Metrology, and related applications:
  - Equipment health monitoring (anomaly detection, alarm, decision/disposition)
  - Equipment / tool digital fingerprinting and matching
  - Feedback and Feedforward compensation and control
  - Supervised/Unsupervised Predictive Models for anomaly detection, virtual metrology, predictive maintenance, yield, parametric disposition, and advanced process control



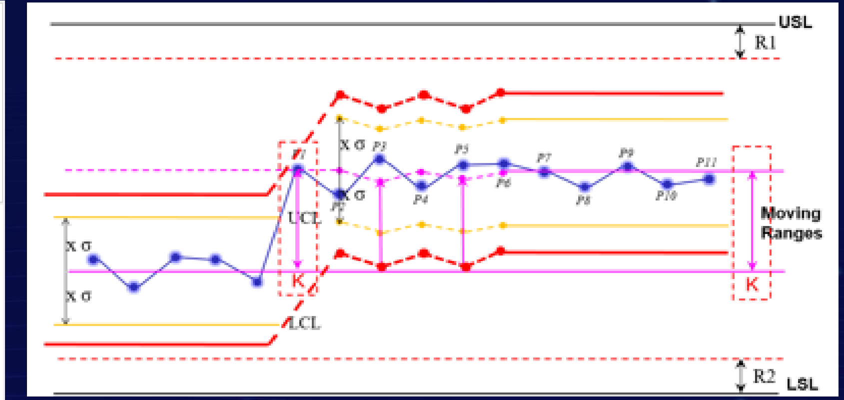
Process Control AI Applications – Modern APC

- Traditional Machine & Deep Learning: APC – Advanced Process Control, FDC – Fault Detection & Classification, SPC – Statistical Process Control, R2R – Run to Run control, VM – Virtual Metrology, and related applications:
  - ❖ Equipment health monitoring (anomaly detection, alarm, decision/disposition)
  - ❖ Equipment / tool digital fingerprinting and matching
  - ❖ Feedback and Feedforward compensation and control
  - ❖ Supervised/Unsupervised Predictive Models for anomaly detection, virtual metrology, predictive maintenance, yield, parametric disposition, and advanced process control

Tool Matching

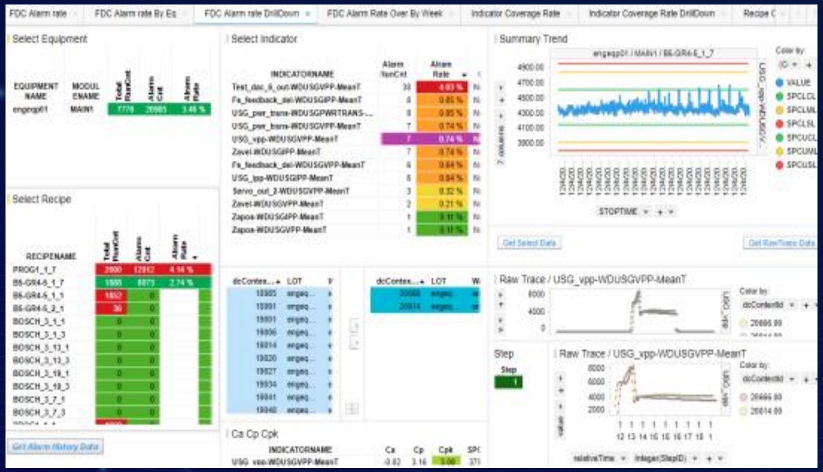


SPC Alarm & Detection



Automatically adjust limits after PM events & drive OCAPs

Health Monitoring

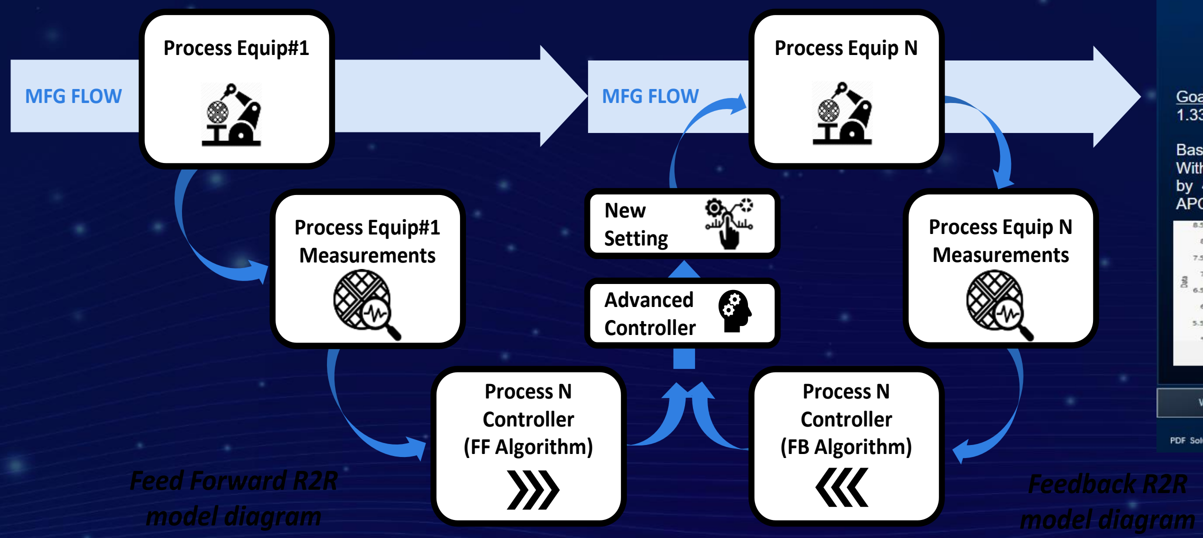


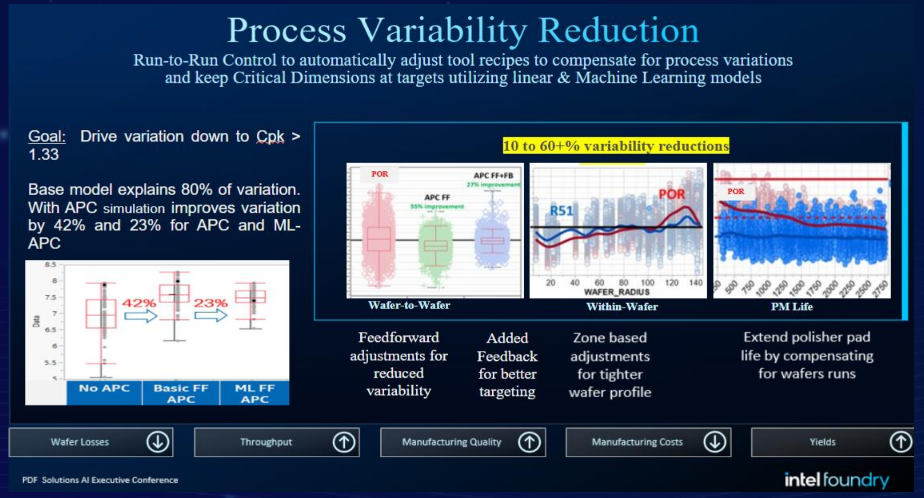
Process Control AI Applications – R2R Control / Feedback

- Traditional Machine & Deep Learning: APC – Advanced Process Control, FDC – Fault Detection & Classification, SPC – Statistical Process Control, R2R – Run to Run control, VM – Virtual Metrology, and related applications:
  - Equipment health monitoring (anomaly detection, alarm, decision/disposition)
  - Equipment / tool digital fingerprinting and matching
  - Feedback and Feedforward compensation and control
  - Supervised/Unsupervised Predictive Models for anomaly detection, virtual metrology, predictive maintenance, yield, parametric disposition, and advanced process control

1. **R2R control** variable process parameter adjustments are made via feedback (FB), feed forward (FF), or FF/FB combination models.
2. **Variable process parameter setting** (recipe setpoint) updates from on-going analysis of historical process and metrology.

R2R control model diagram





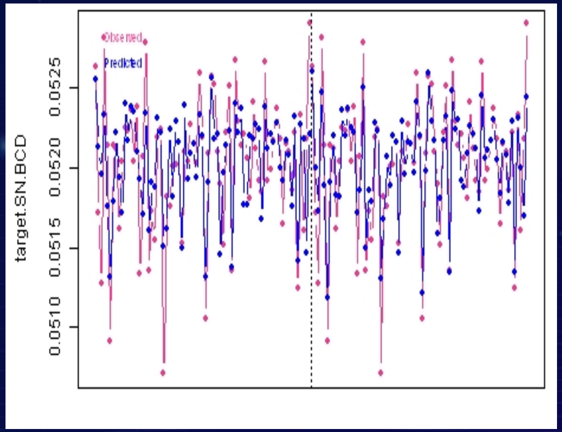
Ref: Aziz Safa, Intel Corporation. Analytics and AI Helping Transformation PDF AI Executive Conference, December 2024

Process Control AI Applications – Predictive Modeling

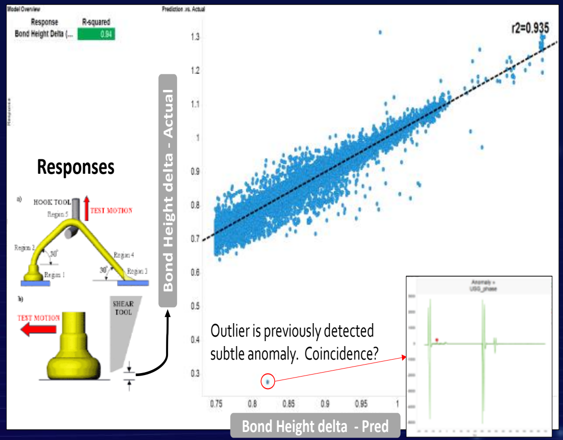
- Traditional Machine & Deep Learning: APC – Advanced Process Control, FDC – Fault Detection & Classification, SPC – Statistical Process Control, R2R – Run to Run control, VM – Virtual Metrology, and related applications:
  - Equipment health monitoring (anomaly detection, alarm, decision/disposition)
  - Equipment / tool digital fingerprinting and matching
  - Feedback and Feedforward compensation and control
  - Supervised/Unsupervised Predictive Models for anomaly detection, virtual metrology, predictive maintenance, yield, parametric disposition, and advanced process control

Unsupervised - Anomaly Detection

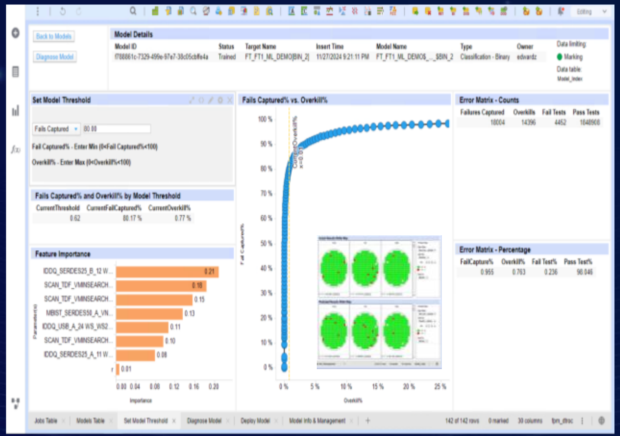
Supervised - Virtual Metrology (CD)



Supervised - Virtual Metrology (WB)



Supervised – Predictive Yield MV Model



Operations Management (scheduling, dispatching, optimization)

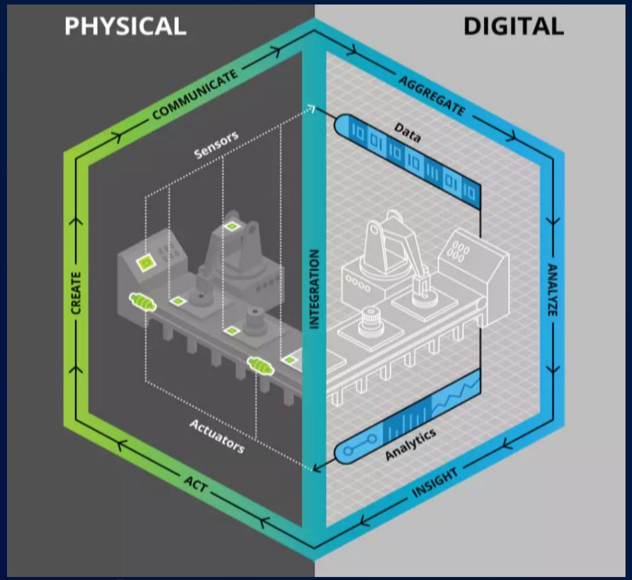
- Smart Manufacturing optimization:
  - **Factory Planning**
    - Capacity/starts planning
    - Forecasting
  - **Scheduling**
    - Commitment plan
    - Maximize thru-put
    - Minimize violations
  - **Dispatching**
    - Commitment execution
    - Maximize thru-put
    - Minimize violations

Factory Planning, Scheduling & Dispatching Solution Scope

SOLUTION	INPUT	OBJECTIVES	OUTPUT	FREQUENCY	APPLIED SOLUTIONS
Factory Planning	• Demand (loading plan) • Planning data (capacity, UPH, CT, PM schedule, etc.) • Material availability and plan	• Determine capacity/wafer starts plan • Forecasting when demand will be met	• Capacity plan • Wafer start plan • Wafer output plan	Weekly or every few days	Multiple solution options •Optimization • Simulation
Scheduling	Current factory status • Equipment • Lots • Process	• Meet due date commitments • Maximize throughput • Minimize queue violations	Lot-to-machine schedules with start/finish times for a shift (8/12 hours for area/Fab)	Every 5/15 minutes	Multiple solution options •Optimization • Simulation
Dispatching	• Lot-to-machine schedules • Real-time factory data	• Meet due date commitments • Maximize throughput • Minimize queue violations	• Dispatch list for each equipment	Real-time	Heuristics

[Integrated solutions to increase semiconductor factory throughput | Applied SmartFactory Solutions]

Digital Twins at the Heart of AI



- Machine Learning and Deep Learning AI models require Digital Twins
  - Component Twins – digital representation of individual parts
  - Process Twins – digital representation of entire production facilities
  - Enterprise Twins – digital representation of entire manufacturing life cycle

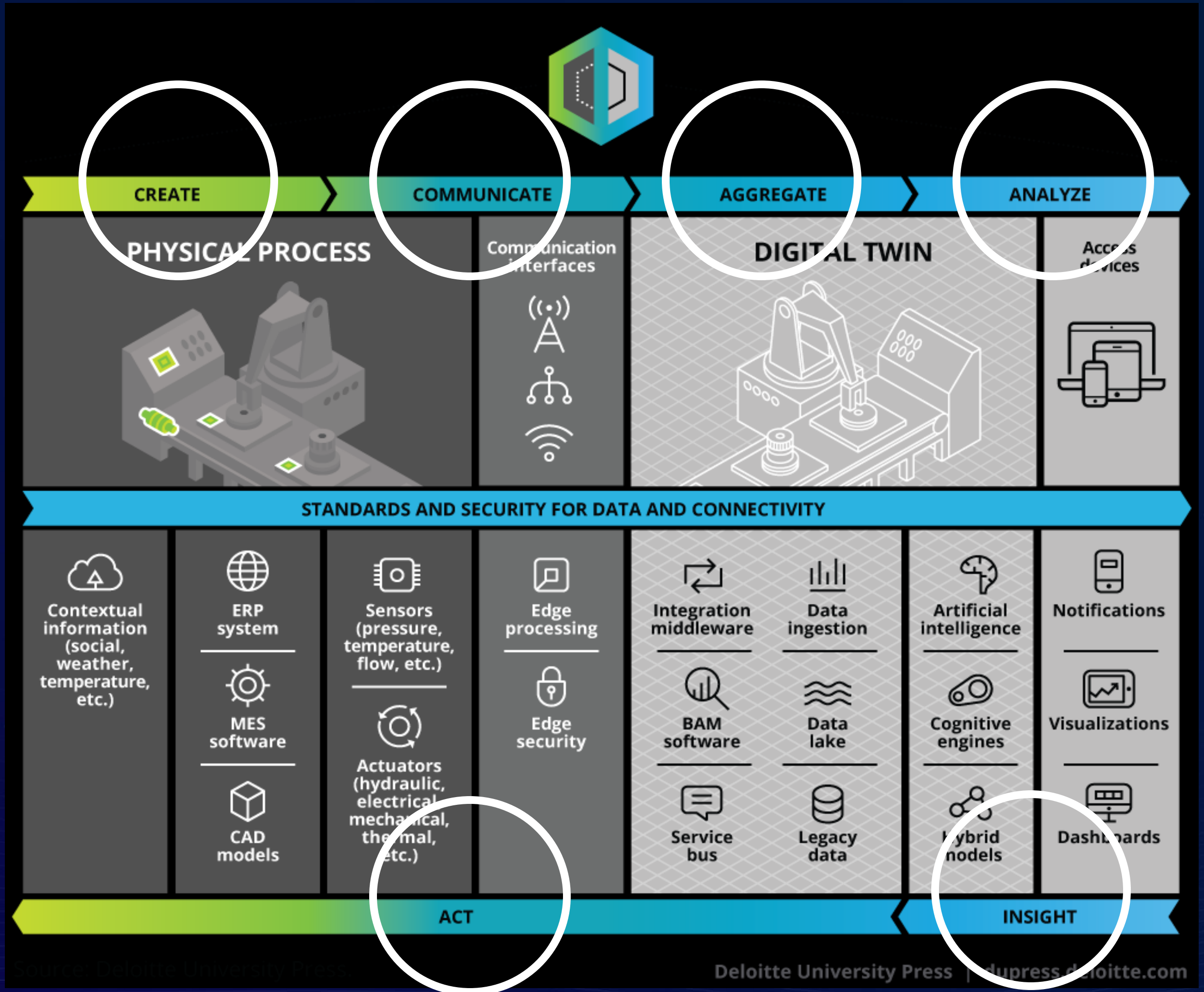
- AI can enhance Digital Twins
  - AI algorithms can automate data collection, integration, and processing, ensuring the digital twin remains accurate and reflects the current state of the physical system.
  - Generative AI can structure inputs and synthesize outputs of digital twins.
  - Digital twins can provide a robust test-and-learn environment for generative AI.

- Some Benefits of AI-Powered Digital Twins
  - **Predictive Maintenance:** Identify potential issues before they occur, reducing downtime and maintenance costs.
  - **Optimized Performance:** Analyze data to identify areas for improvement and optimize operations.
  - **Enhanced Decision-Making:** Provide data-driven insights to support better decision-making.
  - **Improved Efficiency:** Streamline processes and reduce waste.
  - **Risk-Free Testing:** Simulate scenarios and test designs virtually, reducing risks and costs.

Smart Manufacturing applications rely on digital twins

Cyber Physical (Production) Systems

- Key elements on the path to digitizing the IC manufacturing process
  - **Sensors** – Like **FDC**, IoT, Cameras, or cell controllers capturing unique behavior of the physical process
  - **Common platforms** to provide access and link data from the entire manufacturing value chain
  - **Analytics** – Advanced analysis and modeling tools able to be leveraged in manufacturing for things like VM
  - **Knowledge Catalogues & Data models** (digital twins) that allow meaningful and relevant insights to be drawn
  - **Direct CPS/CPPS connections**: Can't ACT if you don't change tool state or decision process
  - **Site-to-site enterprise connectivity**: Must integrate data across the entire supply chain



Source: Deloitte University Press.
https://www2.deloitte.com/insights/us/en/focus/industry-4-0/digital-twin-technology-smart-factory.html#figure-1

AI Scaling Challenges in Semiconductor Manufacturing

Complex Data Model

- From diverse sources: materials, fab, test
- Increasing traceability complexity

Model Maintenance

- Many products / parameters.
- Many machines / sensors
- Large number of models

Complex Interactions

- Large numbers of potential upstream predictors and complex interactions
- Data transformations are often necessary

Diverse Use Cases

- Differing ML pipeline needs mandate customized approaches
- Generalized solution implementation is difficult

Edge Deployment

- Real-time or near real-time processing
- Staging of upstream predictors
- Automated factory and site deployment

Data Drift and Shift

- Need for constant monitoring, alerting and automated actions

Disparate Deployments

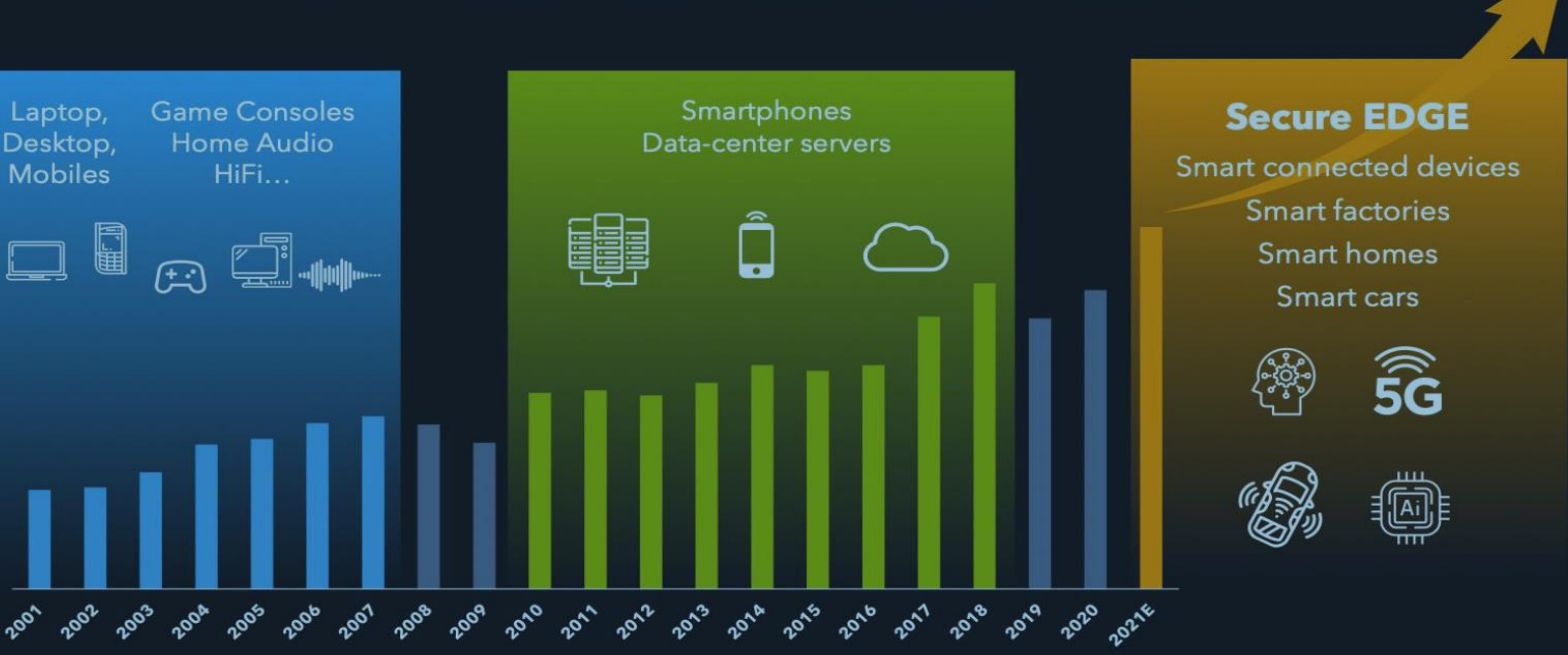
- Deploy to edge, floor server, cloud
- Must be agnostic to different cloud vendors and modeling platforms

Security and IP

- Proprietary algorithms, models, and data, often running on shared systems

RISE OF THE SECURE EDGE

MACROTRENDS DRIVING WAVES OF SEMI GROWTH



Source: IHS, October 2021

New AI Capabilities to Address Challenges

- GenAI – Designed to produce new content such as text, images, videos, audio, and even software code. Uses Machine learning and Deep learning AI (unsupervised, supervised, semi-supervised, or reinforcement)[4]
  - **Downside:** Data dependence for accuracy, very high computational requirements ethical and legal concerns.

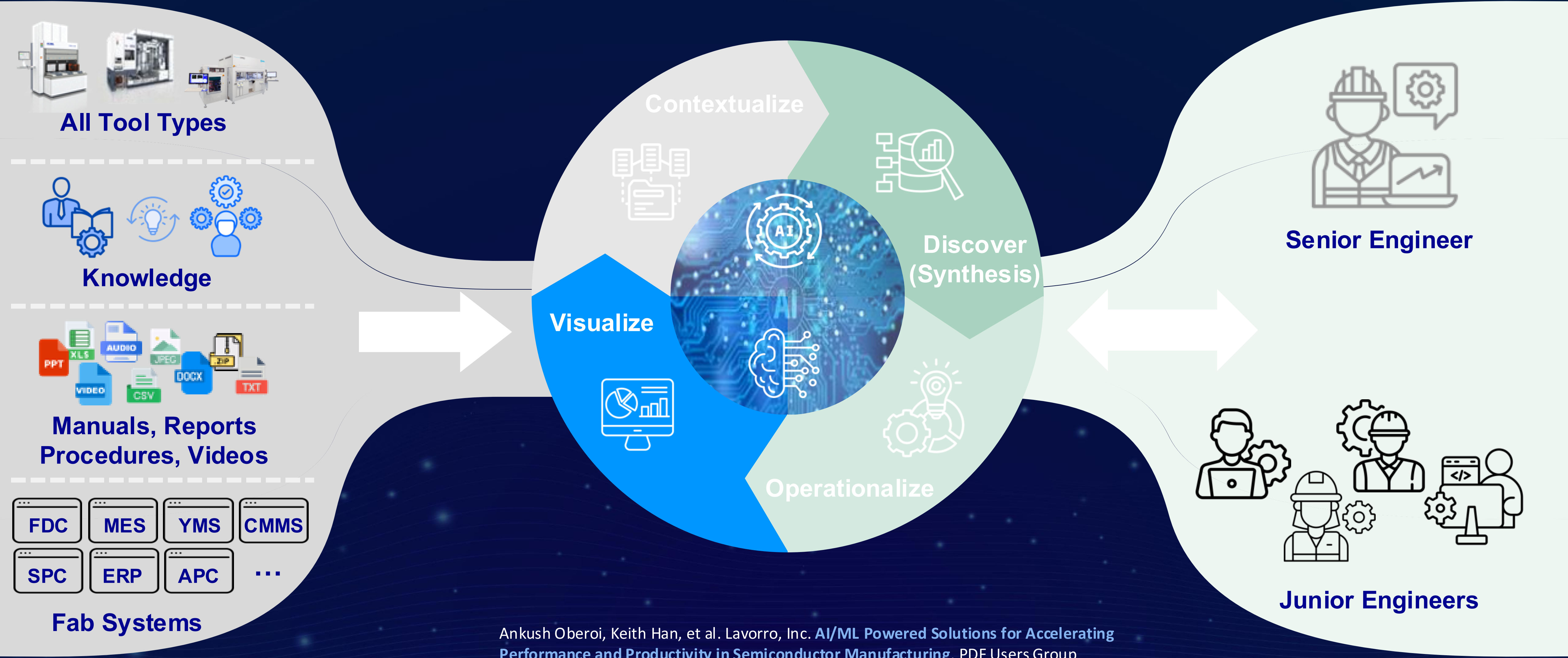
- Large Language Models – Deep learning AI on neural network architectures to process and comprehend human languages. Trained on vast datasets using self-supervised learning techniques. Models intricate patterns and relationships they learn from diverse language data[5]
  - **Downside:** High costs & computational requirements, time intensive, data challenges

- Knowledge Augmented Solutions - Enhances AI systems by integrating knowledge from various sources. Combines Retrieval-Augmented Generation (RAG)'s with knowledge graphs to overcome downsides of GenAI and LLM stand-alone solutions[6].

- Agent AI & Agentic Interaction - AI agents work together in Agentic AI systems to accomplish complex tasks and link Agents together to perform complex tasks

[4] What is Generative AI? | GeeksforGeeks
[5] What is a Large Language Model (LLM) | GeeksforGeeks
[6] Knowledge Augmented Generation (KAG) By Combining RAG with Knowledge Graphs

Knowledge Augmented Solutions For Semiconductor Fabs



Ankush Oberoi, Keith Han, et al. Lavorro, Inc. **AI/ML Powered Solutions for Accelerating Performance and Productivity in Semiconductor Manufacturing**. PDF Users Group Conference, March 2025

AGENTIC AI

systems of multiple AI agents collaborating to achieve complex, multi-step objectives

Learns from outcomes, **shares** episodic/task memory & context **across agents**

Goal-initiated **workflows** that **decompose** and **adapt dynamically**

Source: *Rakesh Gohel*

"AI Agents" – some definitions

1 OpenAI

Agents are systems that independently accomplish tasks on your behalf

2 Google

AI agents are software systems that use AI to pursue goals and complete tasks on behalf of users

3 ANTHROP\C

AI agents are systems where LLMs dynamically direct their own processes and tool usage, maintaining control over how they accomplish tasks

Is agentic AI a repackaging of already-existing tools into an exciting buzzword, or does it represent something new and revolutionary?

The answer is: A little of both…


ELSEVIER

Artificial Intelligence

Volume 117, Issue 2, March 2000, Pages 277-296

Artificial Intelligence

On agent-based software engineering ☆

Nicholas R. Jennings ✉

Show more

+ Add to Mendeley   Share   Cite

https://doi.org/10.1016/S0004-3702(99)00107-1 ↗   Get rights and content ↗

Under an Elsevier user license ↗   ● Open archive

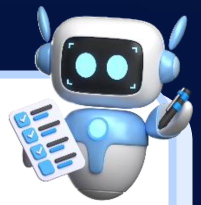
Abstract

Agent-based computing represents an exciting new synthesis both for

Stitching together apps/ functions to do grander things is not new,
so what's new and exciting here?

Source: *https://www.sciencedirect.com/science/article/pii/S0004370299001071*

AI Agents and Agentic AI



AI Agent

RAG retrieves and generates—**AI Agents act**. They perform tasks, run code, call APIs, manage state, and iterate through feedback.

Core Features:

- Task planning & decomposition
- Execution pipelines
- Memory (short & long-term)
- File & API access

Agents shift LLMs from passive responders to **active workflow participants**.

Agentic AI

The next level: **multi-agent systems** with role-based behavior, memory sharing, and communication.

Core Features:

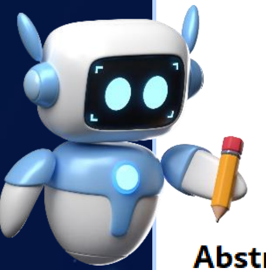
- Collaboration & task delegation
- Modular roles & hierarchies
- Goal-driven planning
- Protocols: MCP (Anthropic), A2A (Google)
- Long-term memory sync & adaptive evolution

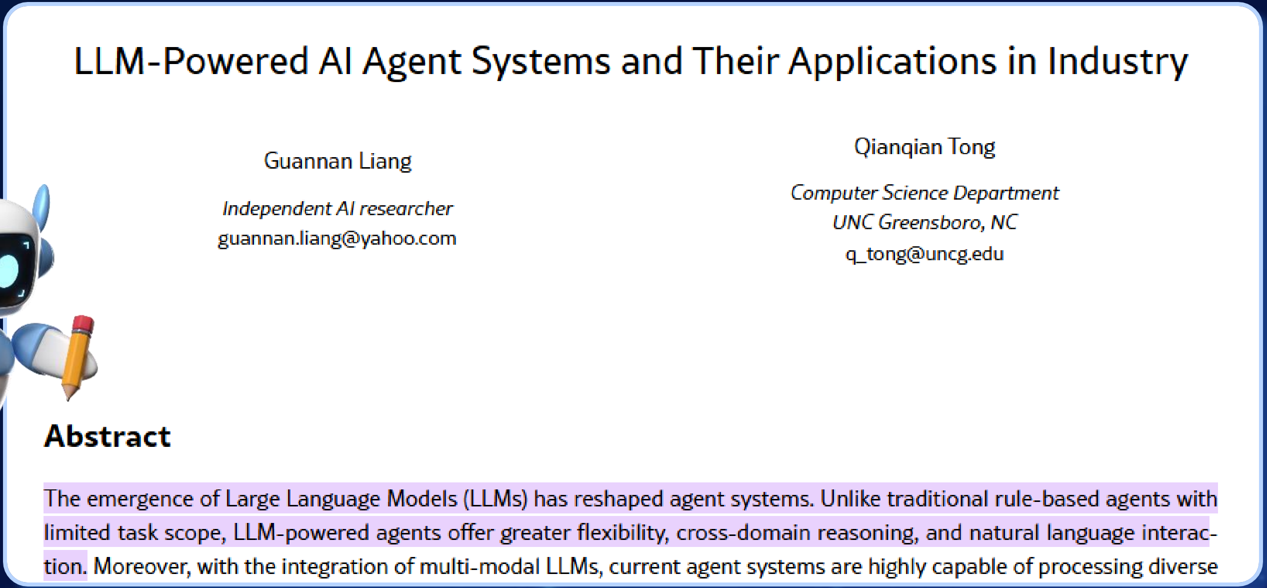
Enables **truly autonomous, collaborative intelligence.**

Source: *Brij Kishore Pandey*

The Right Agents

There is a difference between **LLM** agents and **generic** agents.





LLM-Powered AI Agent Systems and Their Applications in Industry

Guannan Liang
Independent AI researcher
guannan.liang@yahoo.com

Qianqian Tong
Computer Science Department
UNC Greensboro, NC
q_tong@uncg.edu

Abstract

The emergence of Large Language Models (LLMs) has reshaped agent systems. Unlike traditional rule-based agents with limited task scope, LLM-powered agents offer greater flexibility, cross-domain reasoning, and natural language interaction. Moreover, with the integration of multi-modal LLMs, current agent systems are highly capable of processing diverse

Source:
https://arxiv.org/html/2505.16120v1#:~:text=The%20emergence%20of%20Large%20Language,reasoning%2C%20and%20natural%20language%20interaction

In this context, an agent is an **LLM with function calling**.
Function calling = LLM's trained to **provide elements to a function**.

"I have this function which accepts this argument", and the **LLM will give you the arguments** to this function.

LLM's can pick and choose the tools or functions they already have access to and roll up to a workflow.

Agentic Workflows

Agentic workflows = structured instructions executed by agents

We're now applying this more broadly to build custom workflows for complex tasks.

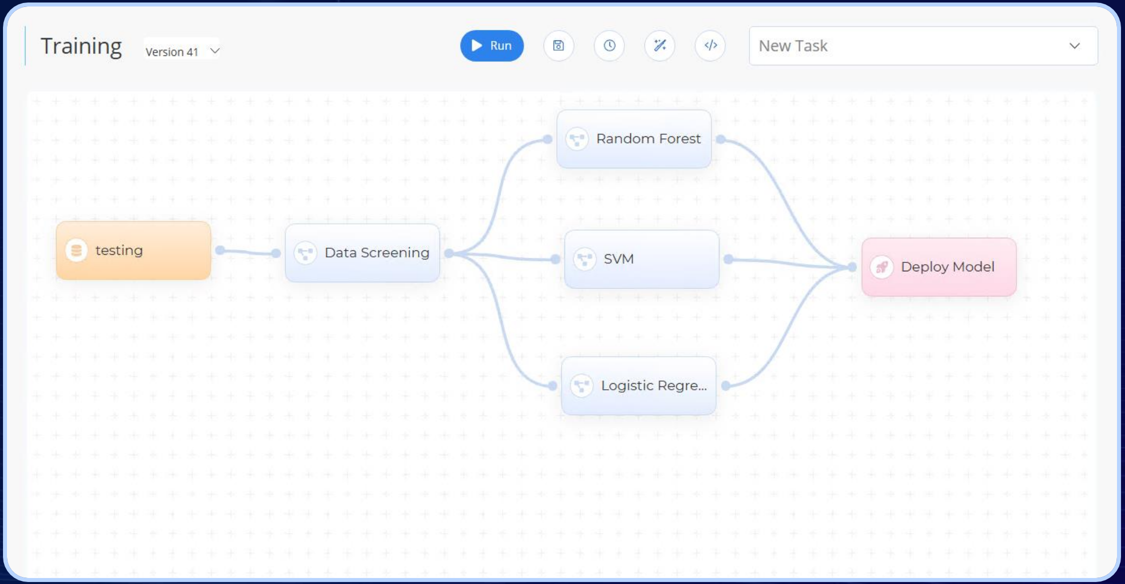
Example

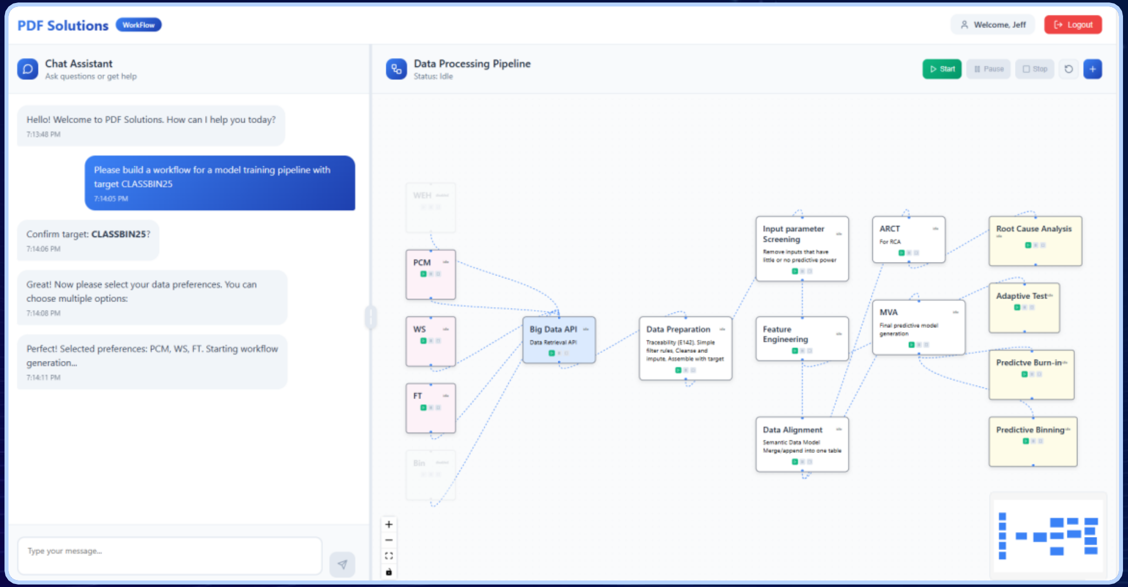
> *"Build a predictive metrology pipeline using X input and Y target."*

Key Differentiator

The agent can **learn and improve** via **feedback loops**—enabling continuous refinement.

Note: Harder to apply this to static knowledge bases.





Model Context Protocol (MCP) Servers

A **standardized way to communicate with LLMs** — led by Anthropic to unify interaction across tools and agents.

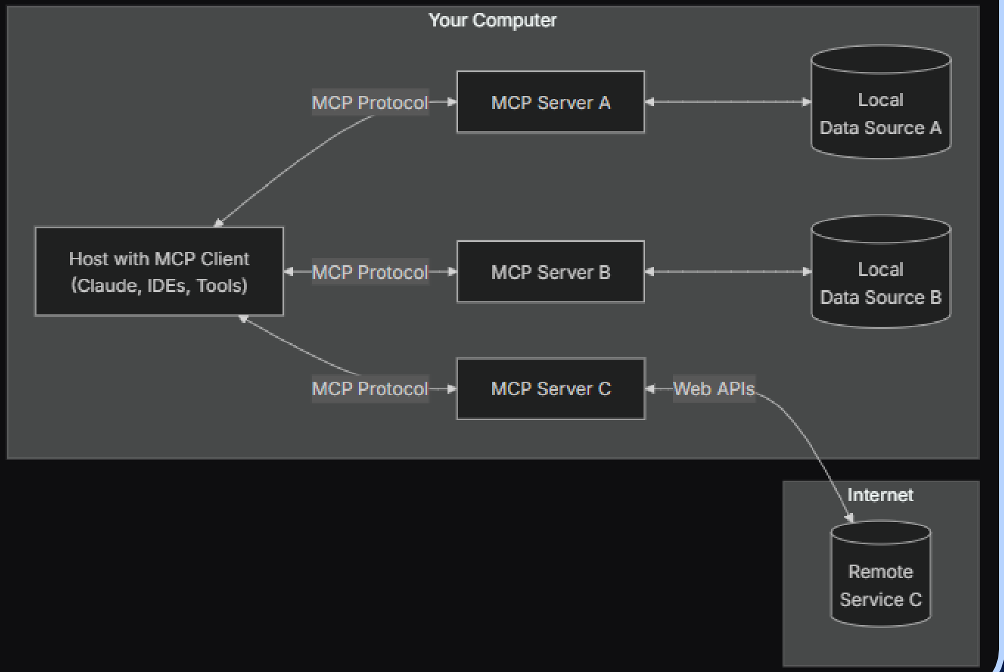
Why?

- encourages **safe, modular integration**
- aims to improve **security** (e.g., reduce prompt injection risks)
- **restricts file access** to defined areas—no full filesystem exposure

A standardized way that our agents can talk to our customers' or other vendors' agents. MCP is the common language between these disparate agents to allow them to communicate and work together.

General architecture

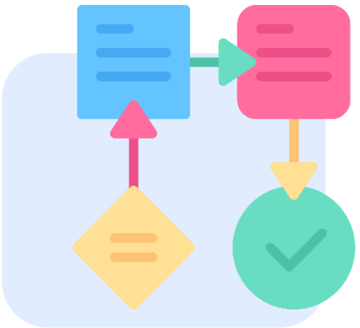
At its core, MCP follows a client-server architecture where a host application can connect to multiple servers:



Key Takeaways

Beyond the hype – Agentic AI is a paradigm shift in our approach to the application of AI systems

AI agents work together in Agentic AI systems to accomplish complex tasks



Agentic AI can be applied to build custom workflows to accomplish these complex tasks

AI Agents can collaborate across industry boundaries to solve some of our biggest challenges

AI Built on Leading Ethics and Privacy Standards

Ensure your data remains confidential

Full transparency and control over your data provided through agentic interaction (limited IP)

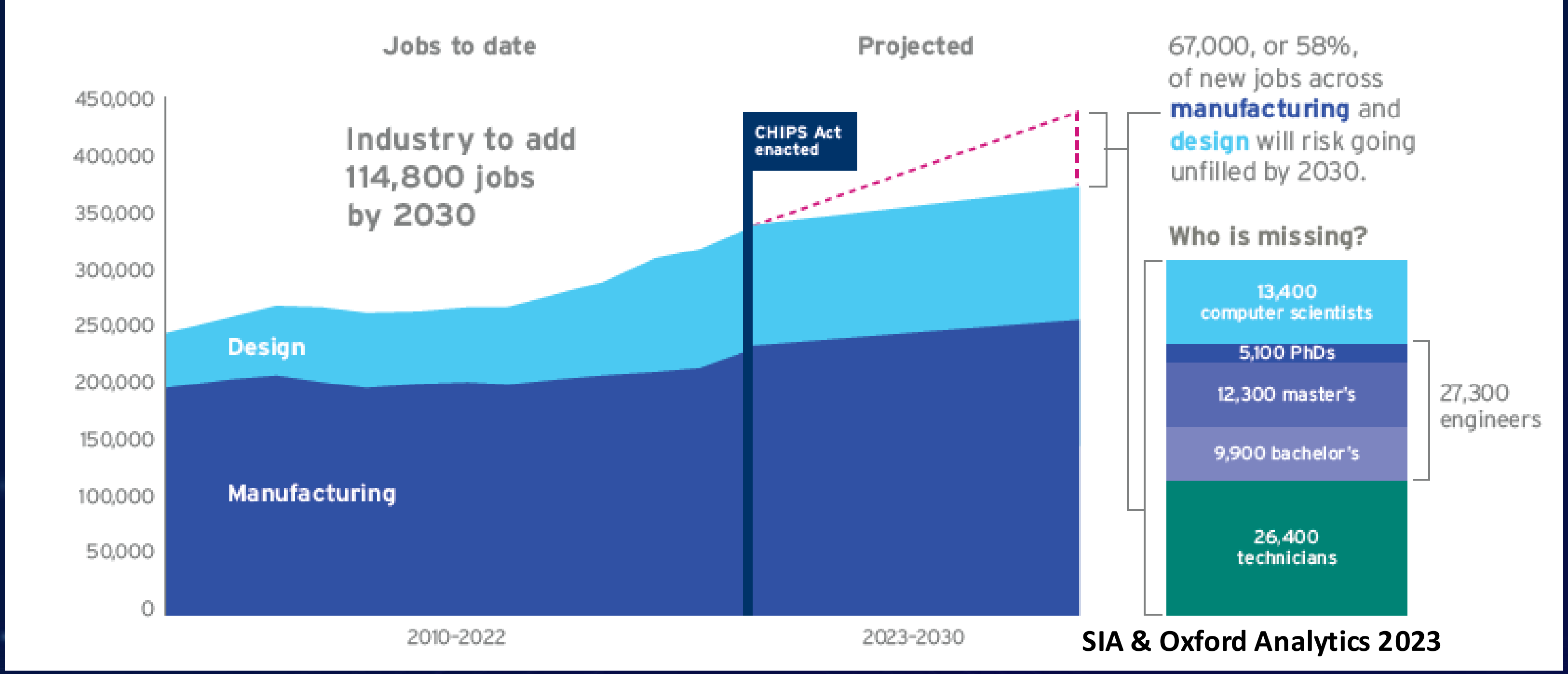
Anonymized data that balances accuracy and maximum privacy protection.

Security and data privacy at the core of every system and process from the start

- Legal framework for data sharing
- Agentic interaction (limited IP sharing)
- Data encryption
- Data anonymization
- Adherence to global standard
- Adherence to privacy regulations
- Security audits
- and more...

Ref: Jon Holt, PDF Solutions, Inc.. Sunil Gandhi, SAP. AI Enterprise Solutions. PDF AI Executive Conference, December 2024

AI is critical in achieving increased employee productivity and breaching the skills gap



Deloitte predicts that more than 1 million additional skilled workers will be needed by 2030

"Highly valued engineers remain bogged down in manual, time consuming fabrication and assembly activities that have the potential to be significantly heightened through automation." – Accenture

"We are definitely in a war for talent,"- Jim Koonmen, EVP at ASML NV

- Develop workforce to use software to bring to scale where foundries use human capital
- Opportunity to leverage AI to revive student interest in Semi jobs
- AI for semi university programs
- Corporate training programs

Conclusions

- In conclusion, while substantial progress in process control has been achieved over the past two decades, the integration of GenAI and LLM-driven knowledge-augmented automated systems marks a transformative leap in manufacturing and supply chain optimization.
- By creating a digital twin and utilizing subject matter experts to refine the model, this system fosters faster decision-making and enhances control mechanisms, resulting in a remarkable 50% improvement in engineering efficiency.
- The system's ability to autonomously generate and deploy and connect highly accurate AI and machine learning models/Agents further drives operational gains, including improved efficiency, reduced cycle times, enhanced yields, and significant cost savings.
- These advancements not only deliver immediate business value but also set the stage for the continued evolution of intelligent, automated manufacturing systems, offering exciting possibilities for future growth and innovation.

Conclusions by ChatGPT

Thank You!