Industrial AI from PILOT to PROFIT

Key Concepts, Success Factors, Use Cases and Market Mechanics

DIE LESEPROBE

2 Key Concepts

2.1 From the Beginning to Today

2.1.1 Foundations (before the 1950s)

In the 1930s, *Alan Turing* envisioned a theoretical device – now known as the *Turing Machine* – capable of performing any computation through predefined rules. Turing also formulated the Imitation Game (later called the *Turing Test*), a conceptual framework for assessing whether a machine can demonstrate behavior indistinguishable from that of a human. He posited that if a machine could convincingly simulate human responses, it might reasonably be considered intelligent. Combined with the invention of electronic computers in the 1940s, this intellectual development marked the transition of artificial intelligence from myth and speculation to tangible scientific possibility.

2.1.2 The Birth of Artificial Intelligence (1950s–1970s)

The formal inception of AI as an academic discipline dates back to the summer of 1956, when a group of prominent researchers convened at Dartmouth College in the United States for a foundational workshop. They theorized that "every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it", and the term *Artificial Intelligence* was introduced during this event.

The initial architects of AI were notably optimistic, hypothesizing that machines would be capable of performing any intellectual task undertaken by humans within one or two decades. Their methodology primarily relied on *symbolic AI*, which entailed programming systems with explicit sets of rules and knowledge – similar to providing comprehensive instructions for each task.

Illustrative early AI programs include the *Logic Theorist*, which could prove mathematical theorems, and ELIZA, which simulated a psychotherapist by identifying keywords and responding with pre-defined phrases. These applications demonstrated the potential for computers to engage in complex logical reasoning.

Despite these early achievements, such systems soon encountered significant limitations. Although effective at solving specific, well-structured problems, they were inadequate for addressing the ambiguous and unpredictable challenges of real-world situations. These systems lacked common-sense reasoning, had limited capacity for experiential learning, and required considerable manual effort to encode all necessary rules. This approach proved unsustainable because it required an exhaustive, rule-based description of the world, which is inherently impractical.

2.1.3 The AI Winters (1970s–1980s)

The initial optimism of the 1950s and 60s began to fade. The grand promises of brilliant machines weren't materializing:

- Early forms of *artificial neural networks*, known as *Perceptrons*, had profound limitations, spotted by thoughtleaders such as Minsky and Papert in the 1970s [2].
- The momentum of *Expert Systems* dried out in the 1980s as researchers realized that intelligence was far more complex than just following a set of rules. It involved intuition, learning from experience, adapting to new situations, and understanding the nuances of human language and the world. The early AI systems simply weren't equipped for this.

Funding for AI research dried up, and public interest waned. These periods are often referred to as the *AI Winters* – a time of reduced enthusiasm and investment.

Despite the winters, important foundational work continued quietly. Researchers explored new ideas that would prove crucial for AI's resurgence.

2.1.4 Learning from Data (1990s–Early 2010s)

As the 1990s dawned, a confluence of factors began to thaw the AI winter. First, computers became exponentially more powerful and cheaper, thanks to *Moore's Law*[1]. This meant machines could process far more information. Second, the Internet began

1 The observation respectively pace setting, that the number of transistors on a microchip doubles roughly every two years.

2.1.5 The Deep Learning Revolution (2010s–2020)

The 2010s witnessed an explosion in AI capabilities, driven primarily by a technique known as *Deep Learning*. This is a specific type of machine learning that uses extensive neural networks – complex, multi-layered mathematical structures loosely inspired by the way neurons connect in the human brain.

What made deep learning so powerful?

- *Even more data:* The Internet and social media generated unprecedented amounts of data for training.
- *Even more computing power:* Graphics Processing Units (GPUs), originally designed for video games, turned out to be perfect for the massive parallel computations needed for deep learning.
- *Improved algorithms:* Researchers developed smarter ways to train these deep networks.

The results were astonishing:

- *Image Recognition:* Deep learning systems became incredibly accurate at identifying objects, faces, and scenes in images, often surpassing human performance in specific tasks.
- *Natural Language Processing:* Machines started to genuinely understand and generate human language, leading to better translation tools, chatbots, and voice assistants like *Siri* and *Alexa*.
- *Speech Recognition:* Converting spoken words into text became remarkably reliable.

In autumn 2012, a significant advancement occurred at a relatively unknown computer vision competition. The *ImageNet Large Scale Visual Recognition Challenge* consistently posed substantial challenges for researchers, who competed to train machines to identify everyday objects such as cats, cars, and chairs across millions of images. Although accuracy had gradually improved, progress was incremental until an unexpected breakthrough emerged. A team led by Geoffrey Hinton and his students, Alex Krizhevsky and Ilya Sutskever, introduced a deep neural network – later known as AlexNet – that dramatically outperformed its competitors by halving the error rate.

For the first time, an automated system achieved image recognition capabilities approaching human reliability. This innovation was highly regarded within academic circles and garnered strong interest from industry stakeholders due to its commercial potential. The application of deep learning quickly expanded: factories transitioned from rigid, rule-based vision systems – which often failed under varying conditions – to models that learned directly from extensive product datasets, thereby improving defect detection efficiency and accuracy.

The adoption of these technologies extended to logistics, where organizations trained AI systems to scan packages, track inventory, and minimize errors in warehouse operations. The medical sector followed suit, with radiologists leveraging *Convolutional Neural Network*-based tools to detect early indicators of disease in imaging studies. Agricultural professionals also benefited, using drones and camera-equipped AI platforms to identify plant health issues and promptly mitigate crop losses.

John Deere's Autonomous Farm Vision

John Deere's journey toward precision agriculture began with the launch of its *MyJohnDeere platform* in 2012, which enabled farmers to collect and manage data from their connected equipment. This laid the foundation for the *autonomous farm vision* unveiled around 2013. A significant acceleration came in 2017 with the acquisition of a Silicon Valley startup specializing in computer vision and machine learning. This expertise enabled the development of advanced solutions, such as the *See & Spray* system, which uses smart cameras and AI to distinguish between crops and weeds, enabling targeted spraying and a significant reduction in herbicide use.

Another iconic moment arrived in 2016 when Google's *AlphaGo*, a deep learning system, defeated the world's top Go player, Lee Sedol. Go is an ancient board game far more complex than chess, requiring intuition and strategic depth. AlphaGo's victory showed that AI could master tasks that seemed uniquely human.

The *Industrial Internet of Things* (IIoT) emerged from the broader IoT concept, applying network connectivity and data exchange to industrial processes and equipment [4]. The early 2010s saw a surge of interest as major industrial companies, aiming to build digital ecosystems for factories and supply chains, developed their own platforms to monetize on the industrial "Big Data" hype. General Electric, a pioneer in this area, introduced its *Predix* platform, positioning it as an open-source operating system for the Industrial Internet [5]. Meanwhile, German industrial giant Siemens launched *MindSphere*, a cloud-based IIoT platform for its own machinery and third-party assets. European companies, including German machine builders like DMG MORI and Dürr, partnered with Software AG to form *Adamos*, a collaborative platform designed specifically for the mechanical and plant engineering sector. Around the same time, American software company PTC made significant acquisitions, including ThingWorx, to build its own IIoT platform focused on rapid application development and connectivity.

Despite the initial hype and significant investment from these major players, the IIoT failed to meet its projected growth, and adoption was far slower than anticipated. This was primarily due to a confluence of technological, cultural, and economic challenges.

From a technological standpoint, integrating new IIoT solutions with legacy operational technology proved to be difficult and costly. Older machines were never designed to be "smart", leading to compatibility issues and a lack of standardized protocols for data communication. Ultimately, this led to data format initiatives such as OPC UA (see Section 9.2.1).

Furthermore, data protection was a significant concern. Connecting sensitive industrial systems to the Internet increased the risk of cyberattacks and the loss of intellectual property and trade secrets.

Beyond technology, other factors contributed to the slow adoption: The high upfront investment costs for hardware, software, and implementation were a significant barrier for many companies, who were also skeptical about the return on investment. The perceived benefits of IIoT – such as predictive maintenance and increased efficiency – were often difficult to quantify in advance (see also Section 5.4).

In this context of applying AI to real-world, high-stakes problems, the term *Industrial AI* began to gain prominence. The concept was notably championed and defined by Professor Jay Lee, who distinguished it from general AI by its focus on industrial systems. He emphasized that Industrial AI required not only predictive accuracy but also extreme reliability, safety, and deep integration with domain-specific physics and engineering knowledge [6].

2.1.6 The Era of Transformer Models (2020+)

More recently, we've entered the era of *Generative AI* (GenAI). Based on an algorithmic class called *Transformers*, systems like ChatGPT can generate coherent, creative text, write code, and even compose poetry. Tools like DALL·E and Midjourney can create stunning, original images from simple text descriptions. These systems don't just analyze data: They create entirely new content, in many cases undistinguishable from content created by human authors.

Launched in late 2022, *ChatGPT* by OpenAI achieved a historic milestone, reaching one million users in just five days, far outpacing the growth rates of previous tech giants such as Instagram and Netflix. This momentum continued, reaching 100 million monthly active users within two months and making ChatGPT the fastest-growing consumer application at the time. Its rapid adoption highlighted a global fascination with generative AI, as the platform quickly transitioned from a viral experiment to an essential tool for millions.

In January 2025, the "DeepSeek moment" sent ripples through the global AI landscape, as a team of leading Chinese algorithmic experts demonstrated that sophisticated model architecture and innovative training approaches could rival or surpass models built primarily on brute-force computational power.

Part II: Adoption Hurdles

3
Why Industry is Harder

Following the rapid adoption of GenAI technology in 2024, there is limited evidence of its impact on business. A recent study by MIT's Nanda initiative reports that 95% of organizations are not realizing returns from their GenAI investments. One reason is the well-documented stalling of use cases during the Proof of Concept (PoC) phase, due to a lack of capabilities and suitable landing zones for production deployment. Another, perhaps less explored, reason is the challenge of deciding between custom AI development and the purchase of third-party products. Nanda's report indicates that externally sourced systems have a 67% higher success rate compared to internally developed tools [56].

What about analytical AI applications in industry? Research indicates that 70% of predictive maintenance programs fail to achieve their objectives, primarily because they cannot integrate the solution into daily workflows [57]. And of course, the specifics of data: While manufacturers collect vast amounts of sensor data, they use only 44% of it effectively. The majority remains trapped in historians or siloed machines, inaccessible for real-time analysis, according to a Rockwell Automation study [58].

3.1 Industry Among Industries

The industrial world is fundamentally different from the data-native environments of social media or e-commerce. These differences create noteworthy barriers to rapid AI deployment:

High-Stakes Environment: The cost of failure is the primary differentiator. If a consumer AI recommends the wrong song, the consequences are trivial. If an Industrial AI miscalculates a parameter in a chemical reactor or fails to detect a flaw in an aircraft engine, the consequences can be catastrophic, involving financial ruin, environmental damage, and loss of life.

Physics and Physical Assets: Industrial systems are governed by the laws of physics, not just by data correlations. AI models must respect these physical constraints to be credible and safe. Furthermore, these systems involve massive, expensive physical assets with lifespans measured in decades, rather than the months or years of a software application.

Brownfield vs. Greenfield: Most industrial environments are brownfield, a complex patchwork of legacy equipment, proprietary control systems, and outdated data protocols from dozens of vendors over many years. This contrasts sharply with the greenfield digital world, where systems are designed from the ground up with modern, standardized data infrastructure.

Beyond shop-floor environments, this also affects PLM, CRM, ERP, and MES environments: wherever stringent Software-as-a-Service strategies are leveraged, rapid AI innovation spreads more quickly than in environments where highly customized systems impede rapid update cycles.

A leading producer of airplane turbines misses out on AI productivity gains as their highly customized PLM system cannot be smoothly updated, hence preventing new AI features from being deployed as they are published by the PLM software vendor.

3.2 Specifics of Sub-Industries

Both the discrete manufacturing and process industries have several sub-industries with varying AI adoption rates. Table 3.1 lists the number and points to the AI adoption drivers:

Table 3.1 AI adoption speed in different sub-industries

Industry	Adoption Speed	Adoption Drivers
Pharma	High	▪ High R&D investment ▪ Large biological and chemical datasets ▪ High-value drug discovery use cases
Automotive	High	▪ High digital maturity in R&D and design ▪ Investment in autonomous vehicles ▪ High costs of recalls
Aerospace & Defense	High	▪ Critical need for reliability and simulation ▪ Complex engineering challenges ▪ High government R&D funding

Industry	Adoption Speed	Adoption Drivers
Consumer Packaged Goods	Medium	■ High volume of sales/transactional data ■ Demand forecasting ■ Personalized marketing and retail analytics
Food & Beverage	Medium	■ Supply chain optimization ■ Production scheduling ■ Quality control and safety monitoring, high costs of recalls
Chemicals	Medium	■ Process optimization and safety ■ Asset performance management ■ High energy costs ■ Complex simulation needs.
Machine Tool Building	Low–Medium	■ Smart factory initiatives ■ Predictive maintenance ■ Automation of complex production lines.

3.3 Potential by Function

According to a McKinsey study, GenAI has substantial economic potential across enterprise functions, with an estimated annual value creation of US$ 2.6–4.4 trillion globally across 63 analyzed use cases. Approximately 75% of GenAI's value creation is said to be concentrated within four key areas [59] (see Figure 3.1).

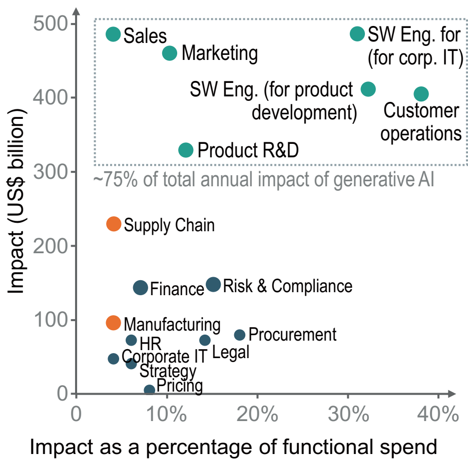

Figure 3.1
Economic potential by enterprise function [59]

4
Making AI Industrial-grade

Building AI solutions for industrial environments requires a far higher bar for robustness, reliability, and longevity than typical consumer or enterprise applications. Industrial-grade AI must be designed to withstand not only technical challenges but also the operational, organizational, and regulatory realities of design, engineering, and production environments.

4.1 Model Level

At the core there are the models themselves. Industrial AI models must be robust– able to cope with shifts in input data distributions without breaking down.

In the context of generative AI, robustness is defined by minimizing hallucinations, as inaccurate outputs can lead to significant or even hazardous outcomes. Hallucinations also tend to impact on acceptance among user groups. Based on our observations, AI models achieving 80% accuracy often generate initial enthusiasm at the management level and are sufficient to proceed to lab testing. However, for genuine user adoption in industrial environments – particularly within engineering cultures – a model typically needs to achieve 95% accuracy.

An engineering team using a third-party LLM to generate PLC code noticed that code produced for identical prompts varied across days, confusing the automation engineers who used the solution. Implementing more than 1,500 test cases that run nightly helps them identify differences and adjust the system prompt to ensure consistent output for users.

Ideally, models are highly capable of generalization and can be transferred across similar machines, processes, or sites without extensive retraining.

An AI model for vibration analysis detects anomalies in pump operation, such as a pump cavity. However, the vibration anomalies would appear in different frequency bands depending on the pump base material (concrete, wood, or steel). In its early versions, the AI model did not generalize well enough to scale as a product. It was only when the model was enriched with a preprocessing function that used noise-filtering algorithms from hearing aid technology that the anomaly detection model became independent of pump location.

Overfitting is a widely recognized issue in AI model training, especially in deep learning. It happens when a model fits the training data – and even its noise – too closely, creating an overly intricate hypothesis that matches the training set exactly but does not work well with unfamiliar data. This results in high variance and weak performance on validation or test sets. The model ultimately "memorizes" examples rather than learning core patterns. Overfitting often stems from using a model that is too complex for the available data or from training it for too many iterations. Increasing the dataset size and employing regularization methods that discourage complex models can help address this problem. A focus on high-quality data is important, as is urging data scientists to move beyond the "easy wins" of overfitting and to adopt a data-centric mindset [60].

AI model performance, when continuously evaluated in an industrial environment, can become a financial liability when AI-driven services are sold based on outcomes or when customers request model performance guarantees. One important mechanism for managing financial liabilities (e. g., those of an AI startup) and associated risks is AI model performance insurance.

Solution Brief 4-1: Insurance against AI Model Liabilities (Munich Re)

Challenge

- Customers lack trust in "black box" AI algorithms and their consistency.
- Uncertainty regarding Return on Investment creates hesitation to invest.
- Companies fear liabilities from AI errors such as hallucinations or IP infringement.
- Risk of financial losses or business interruption due to model underperformance.

Solution aiSure™

- Munich Re evaluates the quality and performance of the specific AI model.
- Provides a performance guarantee backed by insurance to the AI vendor or user.
- Indemnifies clients for financial losses or legal liabilities if the AI errs.
- Covers prediction errors, data drift, and unexpected deviations in behavior

Benefits

- De-risks the purchase decision, allowing for faster AI adoption.
- Protects the balance sheet by transferring financial risk to Munich Re.
- Builds trust with stakeholders and clients through validated model performance.
- Enables companies to scale AI usage without fear of uncapped downside.

Solution Brief 4-1: Insurance against AI Model Liabilities (Munich Re)

Customers: FUGU, TWAICE

Provider: Munich Re, Munich, Germany; *munichre.com/insure-ai*

4.2 Execution Environment Level

Beyond the model, execution must be hardened for industrial conditions. This includes tight integration with cybersecurity frameworks, with every IT/OT service minimized to reduce attack surfaces.

The vast majority of AI workloads are IT workloads, in the sense that they introduce attack vectors that stability-focused OT environments try to avoid at all costs. Particularly when working with industrial AI startups, this can lead to a situation in which ten innovation partners put forward ten different IT stacks and ten different levels of maturity in terms of how stringent these stacks are (security), how they are monitored, patched, and backed up.

A chemical manufacturer learned the importance of execution hardening the hard way. A predictive quality AI system was connected directly to the plant network without proper patching. When malware hit the broader IT environment, the unprotected AI system went offline, halting production for several hours. The incident led to a company-wide overhaul of cybersecurity integration.

Solution Brief 4-2: Industrial Edge Platform (Siemens)

Challenge

- Critical machine data remains trapped on the shop floor without easy cloud access.
- Real-time processing is impossible when relying solely on slow, distant cloud connections.
- Managing software and security updates across thousands of decentralized industrial devices is difficult.

Solution "Industrial Edge"

- Processes high-frequency data locally at the machine level with support for OT network protocols, MQTT, and PLC connectivity.
- Provides a global hub to deploy, update, and monitor apps across all devices.
- Supports diverse industrial applications and Docker-based software for flexible, customized functionality, based on a hardened Linux stack ('Industrial OS').

Solution Brief 4-2: Industrial Edge Platform (Siemens) *(continued)*
Benefits ▪ Protects industrial assets with automated security patches and secure, isolated data processing. ▪ Bridges need for innovation (e. g., startup solutions) and stability in corporate large-scale up to global OT environments.
Customers: BMW, Festo, Volkswagen
Provider: Siemens, Berlin/Munich, Germany; *siemens.com/industrial-edge*

When the execution environment is cloud-based, common hardening methods include API monitoring (network traffic to and from the AI solution), network traffic monitoring between cloud components, and robust identity and access management, including session-based usage tokens.

4.3 Hardware Considerations

Industrial AI hardware presents its own set of challenges, some related to technical performance, but many more to supply chain resilience and other factors. It is fair to say that Proof of Concept projects hardly embrace the hardware dimension in their feasibility focus, which is one of several reasons for expensive "Lab to Reality" gaps:

- *AI chip power consumption:* The sheer amount of power required by some AI chips imposes challenges, specifically when the AI system is embedded into battery-driven, mobile systems such as AGVs or AMRs.
- *AI chip heat dissipation:* As a function of power consumption, AI chip heat must be safely dissipated. Does this require a fan, or can the edge device operate with a fanless design, which offers many advantages in dusty or hygienic shop-floor environments?
- *AI chip floating point calculations:* Different AI chips show small differences in their calculation routines, which means their floating point numbers differ in terms of the last decimal digits. This causes the AI model's outputs to differ depending on the hardware on which it is executed.
- *AI chip export control restrictions:* The rising geopolitical tensions and AI technology in the middle of an international powerplay have already led to restrictions on AI chip exports. Developing globally available products (e. g., AGVs) already requires regional product variants with different AI chips.

Figure 4.1 Siemens Copilot user interface design

Shibaura Machine collaborated with *Spur Insights* in Switzerland to develop the LEO AI Assistant to support operators of its injection molding machines. LEO has been personified as a helpful blue-collar colleague with a very positive impact on user acceptance (see Figure 4.2).

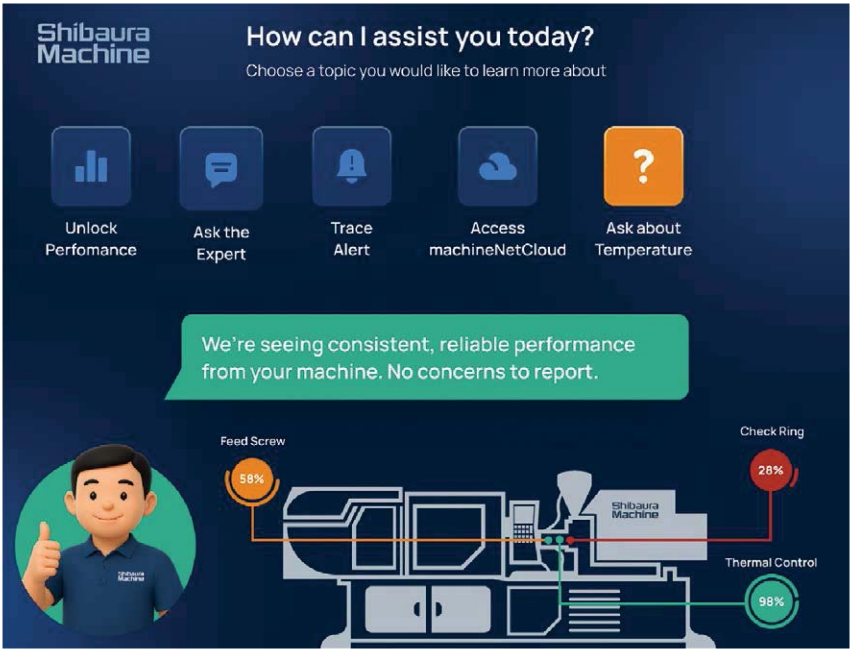

Figure 4.2 Shibaura Machine's LEO AI Assistent

IBM named their powerful AI "Watson", creating the subliminal association that the human in front of the system is equal to the genius Sherlock Holmes and can rely on the helpful services of a junior partner.

- *Avoiding AI jargon:* To ensure the successful launch of industrial AI solutions, developers must bridge the communication gap by replacing data science jargon with the practical language of the shop floor. Terms such as "F1 Score" and "hyperparameter tuning" often alienate operators, creating a perceived complexity barrier that undermines trust. When technical concepts are translated into domain-specific outcomes – such as "Decision Confidence" or "Machine Health" – the AI stops feeling like a foreign intrusion and starts being viewed as a functional tool. Priori-