McKinsey & Company

McKinsey Technology Trends Outlook 2022

Industrializing machine learning

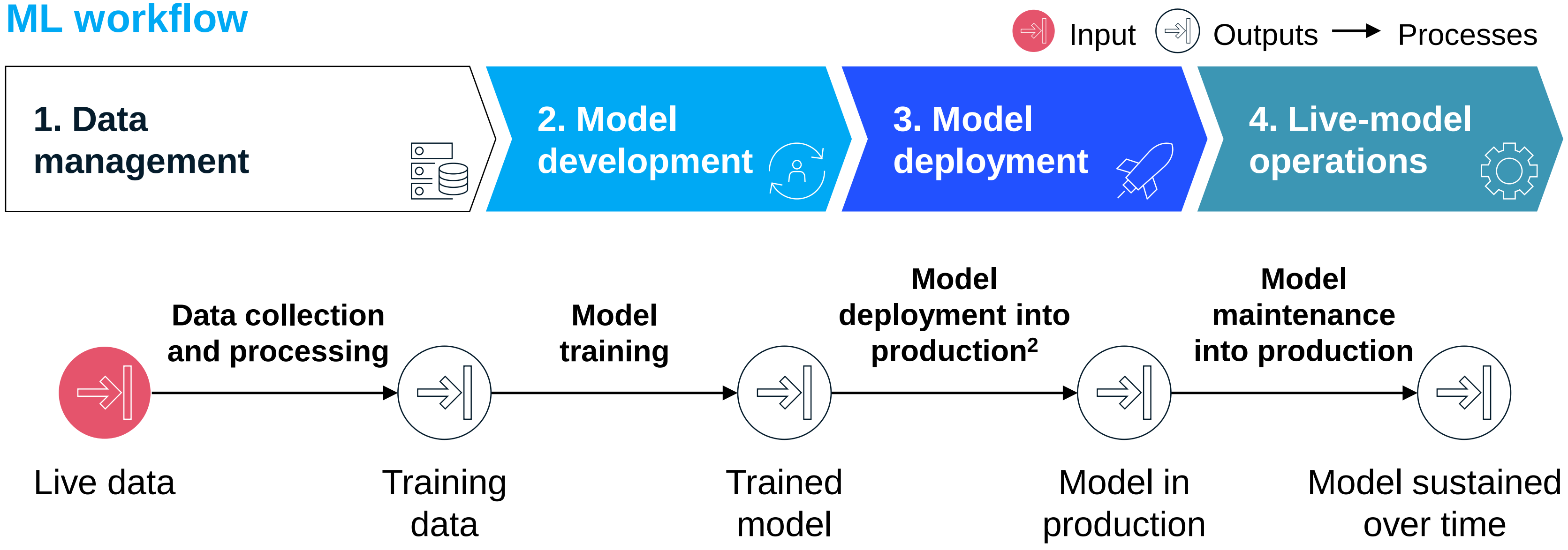
August 2022

What is the trend about?

Machine learning (ML) workflows are the processes that bring AI and ML into production for real-world business use

Solutions industrializing ML provide the **software and hardware technologies to scale ML workflows and ease the development and deployment of ML** for organizations[1]

ML workflow



[1]To differentiate applied AI and industrializing ML, this tech trend refers to the systems that put AI (including its subfields such as ML) into production for real-life business use. Applied AI refers to the real-world business use cases after the technical infrastructure is implemented.
[2]Once performance standards are met.

Source: McKinsey analysis

Future progression

ML production for organizations is **delivered reliably** and **at scale,** featuring:

- deployment scaled across networks
- modular structure with high reuse
- robust monitoring and testing
- automation of common processes
- low maintenance cost, lower risk, higher ROI

Why should leaders pay attention?

Solutions for industrializing ML address the technical challenges that prevent organizations from unlocking the full potential of AI and ML

AI is becoming essential for success …

AI has massive potential across industries

$10T–$15T
global impact potential

AI is disrupting traditional business models

5
of 10 largest global companies did not exist 25 years ago

AI leaders have stronger financial performance

2.5×
greater 5-year total shareholder returns

… but challenges remain

72%

of organizations surveyed have not successfully adopted and scaled[1]

Challenges include:

- Difficult transition from pilots to products
- Model failure in production
- Stalling team productivity
- Limitations in protection against potential risks from unknown variables

[1]McKinsey survey with >1,000 company executives who launched transformations. Organizations get stalled in the pilot phase or during scaling, or they have limited impact despite scale.

Source: “Notes from the AI frontier: Applications and value of deep learning,” McKinsey Global Institute, April 17, 2018; S&P Capital IQ (Oct 2020); McKinsey analysis

Why should leaders pay attention? (continued)

Industrializing ML has potential for impact on all industries by reducing hurdles to develop ML in a reliable manner[1]

Value levers	Impact potential within 1 year[2]
Maintain performance	**~60%** less value erosion 12 months from model deployment because of live-model operations
Accelerate time to value	**~8–10×** less time from proof of concept to production system because of standardization from data management to model deployment
Reduce risk	**100%** of production models integrated into enterprise risk governance and fully auditable because of interoperable systems
Increase productivity	**~30–40%** resource reduction for ML operations with improved automation

[1]Impact also associated with the applied-AI tech trend.
[2]Based on observations from ML operations deployment in 5 large-scale analytics transformations supported by McKinsey.

Source: Expert input; McKinsey analysis

Why are the technologies interesting, compared with what already exists?

This emerging tech stack is moving toward simplicity, scalability, and interoperability across the full ML workflow life cycle

	From	To
Overall ML workflow	**Outdated tools** operated in an **inefficient manner** **Fragmented technology landscape** creating inconsistent standards and **limiting collaboration** between teams	**Tooling optimizing ML workflows** Technology enabling **learning to be shared** and **collaboration** across the business, including technical and nontechnical employees
Data management	**Massive manual effort** for **one-off use** with **no controls** over quality and drift, which can damage overall performance	**Automated** data management for **high-quality data** **Data reuse** across hundreds of solutions with **robust controls**
Model development	**Individualistic, artisan** experimentation **Highly manual** work that needs **refactoring before use**	**Structured and collaborative** development **Solutions assembled from prebuilt components** and tooling with high degree of **automation**
Model deployment	**No control** over what is running in production **Manual and error-prone** deployment with **poor testing and validation**	**Controlled production release** decisions **Model management providing full transparency** on production solutions **Automated, efficient CI/CD[1]** for **test and validation of all releases**
Live-model operations	**Performance degrading,** often undetected, **eroding model value** **Unstable solutions** down for weeks at a time	End-to-end ML **system monitoring** with instant alerts enabling rapid issue resolution

[1]Continuous integration (CI) and continuous delivery (CD).

Source: Expert input; McKinsey analysis

What are the most noteworthy technologies?

Software solutions across the ML workflow

ML workflow

	1. Data management	2. Model development	3. Model deployment	4. Live-model operations
ML subprocesses	Data discovery and creation • collection Data transformation • labeling • validation Data versioning	Feature engineering Model development and optimization • model selection • training and tuning Experimentation and testing	Model registry and management Model testing and validation Model deployment • continuous integration • continuous deployment	Monitoring (eg, infrastructure, data, model, key performance indicators) Model maintenance • continuous improvement Model explainability
Examples of technology solutions	Data platforms for data discovery and extraction Synthetic data Automated data labeling Reproducible data pipelines	Model libraries Centralized repository for experimentation Experiment tracking and model visualization tools Developer environments (notebooks and IDEs[1])	Packaged testing and deployment platforms Version and result tracking	Monitoring of live-model, data pipelines, and related issues Continuous retraining with periodic refreshing of data sets

Plug-and-play end-to-end platforms

[1]Integrated development environments.

Source: Expert input; McKinsey analysis

What are the most noteworthy technologies? (continued)

Hardware solutions for software interconnection and workload optimization

Integrated hardware

Solutions connecting physical hardware chips and software frameworks

Vertically integrated hardware systems

Specialized hardware–software solutions that are tailored to specific ML tasks; service-based examples include access management service to GPUs and data flow as a service

Horizontally integrated hardware

Hardware offering a diverse and broad set of solutions (eg, simplifying use of distributed compute)

Heterogeneous computing

Solutions optimizing computational workloads by allocating different hardware chips based on specific task[1]

Graphical processing units (GPUs)

Hardware useful for linear-algebra-based computations

AI-specific GPUs are being tailored with faster training speeds, faster transfer speeds, and stronger computing power

Tensor processing units (TPUs)

Specialized hardware useful for deep-learning computations and able to handle complex linear algebra (eg, "tensor" or matrix multiplications)

Neuromorphic processing units (NPUs):

Early-stage hardware chip based on brain neural-network architectures with potential impact for low energy consumption

[1]With the end approaching for Moore's law and Dennard's law, where computational growth grows exponentially, new solutions for optimizing computing are being explored. The solutions here are different computer chips that can be leveraged for different tasks.

Source: Expert input; McKinsey analysis

What should leaders consider when engaging with the trend?

Benefits

Accelerated AI adoption due to reduced technical barriers and requirements to implement AI

Improved productivity for technical employees across ML life cycle

Easier collaboration between technical and nontechnical experts on ML model development

Scalability and interoperability leveraging bigger, richer reused data sets

Reduced cost through faster development and deployment, standardized processes, improved technical performance

Improved security and privacy along with reduced risk due to greater standardization and process automation, transparency, and robustness

Risks and uncertainties

Upfront investment and resources for setup, where organizations need ML-savvy talent and processes to build capabilities and accelerate the learning curve across the organization

Dependency on 3rd-party vendors leading development of ML technologies for initial onboarding and continuous support

Fast-developing market, where processes and accountability for maintaining ML solutions have been poorly defined

Increasing regulation and compliance, where legislation can affect ML's development (eg, data governance policies affecting data management solutions)

Increasing need for responsible and trustworthy ML systems to address concerns about ethics, privacy, equity and fairness, explainability, accountability, security, and governance

Source: Expert input; McKinsey analysis

Who has successfully created impact by industrializing ML?

Solutions for industrializing ML have generated value for organization as part of large-scale analytics transformations

	Global pharma company	Asian digital bank	North American retailer
Context	Integrating **personalized recommender system** for **>**50 country and drug combinations **Sales uplift of 10–20%**	Industrializing and scaling AI use cases to become a truly data-driven bank Enabling >**150 use cases** annually	Applying solutions industrializing ML to address scaling pain points Scaled up **end-to-end analytics transformation** from only 5 models in production
Technology used[1]	**Software for model development** **Software for model deployment** **Software for live-model operations**	**Software at all stages** **Hardware at all stages**	**Software for model deployment** **Software for live-model operations**

Impact achieved

Global pharma company	From	To	Asian digital bank	From	To	North American retailer	From	To
Time to market	4 months	1 month	Lead time to production	18 months	3 months	Deployment process	Manual and error prone	Automated and reliable
Code compatibility check	2 weeks	2 hours	Asset reuse	<10%	>50%	Deployment lead time	2 weeks	1 day
Tech setup time	3 days	15 minutes	Enterprise risk coverage	0%	100%	Deployment time	>8 hours	15 minutes

[1]Diverse solutions used according to the technology type and subprocess.

Source: Expert input; McKinsey analysis

What industries could be most affected by the trend?

A **diverse set of stakeholders** across a range of industries are experiencing implications from the industrialization of ML; impact from this trend is most expected in industries where accelerating production of ML application yields a competitive advantage

Industry affected	Examples of impact from tech trend
 **Information technology and electronics**	Designing hardware and software so that devices become more integrated and connected with the natural world (eg, AI models to interpret voice commands, sensors)
 **Telecommunications**	Deployment to aid business functions from marketing and sales (eg, upselling or cross-selling engines) to customer service (eg, call center volume forecasting and predictions) and network optimizations
 **Pharmaceuticals and medical products**	Supporting the development of new drugs, (eg, through exploring relationships between molecules and chemical compounds) and enabling support functions (eg, manufacturing, supply chain optimization) for various medical treatments
 **Aerospace and defense**	Augmenting design and manufacturing processes through optimizations from AI/ML models (eg, AI models to aid in the 3-D simulations for aircraft design, supply chain optimization for manufacturing, security risk management)
 **Automotive and assembly**	Using AI/ML to enhance design and manufacturing processes such as predictive maintenance, automated quality testing, and demand forecasting and to provide customer service features such as navigation
 **Financial services**	Supporting key services in the financial sector including risk management and assisting in many other processes—eg, by detecting credit card fraud
 **Media and entertainment**	Provision of high levels of personalization in media and entertainment experiences (eg, tailored recommendations)

What are some topics of debate related to the trend?

1 Impact of ML industrialization on organizations and technical talent

How can solutions that industrialize ML change organizations, their operating models, and their engineering roles?

The technologies are part of a wider effort in scaling ML operations toward a **modular, automated, monitored life cycle approach** to AI/ML—potential impact:

- **reduce resourcing needs and production times**
- **'democratize'** (ie, use nonspecialized) **data scientists** working horizontally on **most value-add tasks,** assisted by standardized tooling
- reduce technical barriers and enable **closer collaboration with nontechnical SMEs,**[1] offering greater visibility and expanding potential use cases

2 Selection criteria for solution that industrialize ML

How should organizations select which solutions that industrialize ML are most relevant to their needs and strategy?

- **Industry-specific use cases** influence **ML workflows,** varying drastically across risk levels, required data governance, SME relevance, and model complexity
- Potential long-term dependency on 3rd-party vendors means organizations may have **long-term partnerships rather than solution providers,** often making trade-offs between **best-of-breed vs end-to-end/cloud-native ML platforms** and **open-source vs supported enterprise software**

3 Accountability for AI and ML

As the solutions that industrialize ML grow, how can roles of accountability be defined to ensure trustworthy and responsible AI/ML?

- Processes and accountability for maintaining ML solutions are **currently poorly defined,** with **lack of clarity on roles** of responsibility across the ML workflow
- As with applied AI, organizations will have to make **trade-offs on which aspects of trustworthy AI** are priorities for their business, which will have downstream impact on their decisions and interactions with solutions industrializing ML

[1]Small and medium-size enterprises.

Source: Expert input; McKinsey analysis

Additional resources

Knowledge center

[QuantumBlack, AI by McKinsey]

Related reading

Scaling AI like a tech native: The CEO's role

Operationalizing machine learning in processes

Transforming advanced manufacturing through Industry 4.0

Derisking machine learning and artificial intelligence