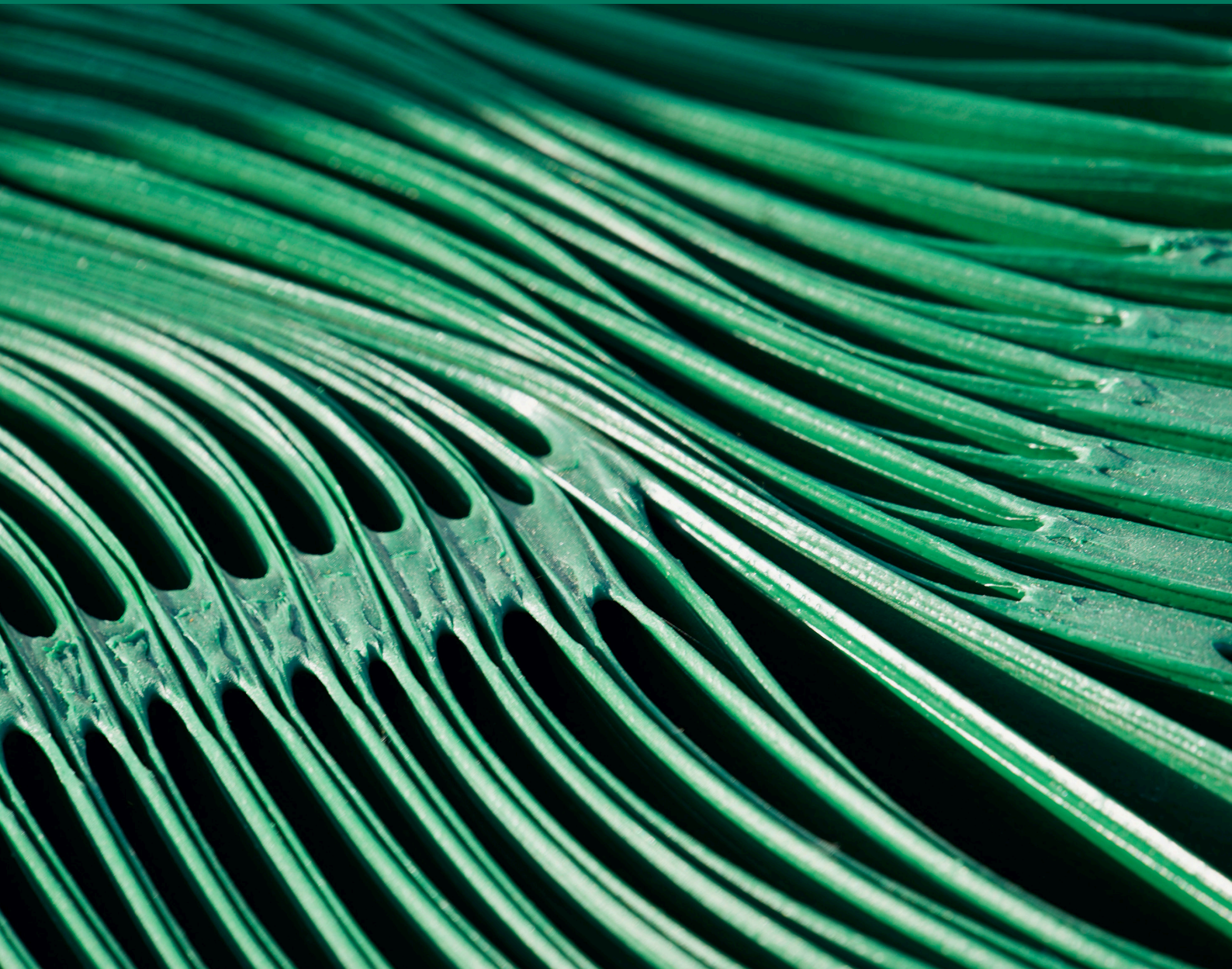


BCG

THE BOSTON CONSULTING GROUP



The Big Leap Toward AI at Scale



The Boston Consulting Group (BCG) is a global management consulting firm and the world's leading advisor on business strategy. We partner with clients from the private, public, and not-for-profit sectors in all regions to identify their highest-value opportunities, address their most critical challenges, and transform their enterprises. Our customized approach combines deep insight into the dynamics of companies and markets with close collaboration at all levels of the client organization. This ensures that our clients achieve sustainable competitive advantage, build more capable organizations, and secure lasting results. Founded in 1963, BCG is a private company with offices in more than 90 cities in 50 countries. For more information, please visit [bcg.com](https://www.bcg.com).

The BCG Henderson Institute is The Boston Consulting Group's internal think tank, dedicated to exploring and developing valuable new insights from business, technology, and science by embracing the powerful technology of ideas. The Institute engages leaders in provocative discussion and experimentation to expand the boundaries of business theory and practice and to translate innovative ideas from within and beyond business. For more ideas and inspiration from the Institute, please visit <https://www.bcg.com/bcg-henderson-institute/thought-leadership-ideas.aspx>.



THE BOSTON CONSULTING GROUP



The Big Leap Toward AI at Scale

Philipp Gerbert, Sukand Ramachandran, Jan-Hinnerk Mohr, and Michael Spira

June 2018

AT A GLANCE

As companies introduce artificial intelligence, they are stumbling up against the “AI paradox”: it is deceptively easy to launch AI pilots but fiendishly hard to reach scale.

A CEO’S GORDIAN KNOT

The way machines are organized is as critical as the way people are organized. Companies need to install end-to-end management platforms that harness the raw power of AI.

IN VENDORS WE MUST TRUST

Companies often depend on vendors who need access to critical company data to build “intelligence.” These relationships can determine competitive advantage.

THE HUMAN TOUCH

AI requires new technical skills (beyond just data science chops) and a heightened level of expertise in change management and strategic workforce planning.

ORGANIZATIONAL BALANCING ACT

AI requires new organizational principles: data governance and AI frameworks at corporate; algorithmic learning centralized within business units; and actions of AI engines decentralized.

AFTER DECADES OF FALSE starts and unfulfilled expectations, artificial intelligence (AI) has now gone mainstream. Companies are successfully applying AI to a wide variety of current and novel processes, products, and services, and this success comes none too soon; AI is essential for business to address the ballooning complexity brought on by digitization and big data.

Visionary executives have started to imagine the potential for implementing AI throughout their companies. For example, BCG estimates that AI could help the top ten banks generate an additional \$150 billion to \$220 billion in annual operating earnings.

As pioneers across industries strive to reap these rewards by scaling up AI, however, they are stumbling against what we call the “AI paradox”: it is deceptively easy to launch AI pilots and achieve powerful results. But it is fiendishly hard to move toward “AI@scale.” All sorts of problems arise, threatening to undercut the AI revolution at its inception.

The paradox is easy to explain but hard to resolve. AI forces business executives to deal simultaneously with technology infrastructure as well as more traditional business issues. The core challenge is a tightening Gordian knot. Typical IT systems consist of data input, a tool, and data output. These systems are relatively easy to modularize, encapsulate, and scale. But AI systems are not so simple. AI algorithms learn by ingesting data—the training data is an integral part of the AI tool and the overall system. This entanglement is manageable during pilots and isolated uses but becomes exponentially more difficult to address as AI systems interact and build upon one another.

What sounds like a mere technical issue has multifaceted implications and challenges for companies. For example, vendor management becomes both more strategic and more complex. For the foreseeable future, vendors will play a large role in deploying AI because they have hired so much AI talent and have even promised to help with entanglement. Companies need to prepare for working with AI vendors in ways that do not put their data at risk or create long-term dependencies but instead strengthen competitive advantage.

People challenges also loom large for AI@scale. On one hand, companies confront a scarcity of AI data scientists and systems engineers; on the other hand, current employees are concerned about interacting with machines and even about losing their jobs.

It is deceptively easy to launch AI pilots and achieve powerful results but fiendishly hard to move toward AI@scale.

Also, the organizational demands of AI require a delicate three-dimensional balancing act. Data governance, core AI expertise, and system management should be centralized. The development of use cases, learning, and training should occur in business units or functions and be managed by agile, cross-functional teams. Finally, AI action remains decentralized—in the marketplace, on the shop floor, or in the field.

Below, we provide a roadmap to systematically resolve the AI paradox. The AI@scale program is a full transformation to a new operating model.

A Systematic Approach to AI@scale

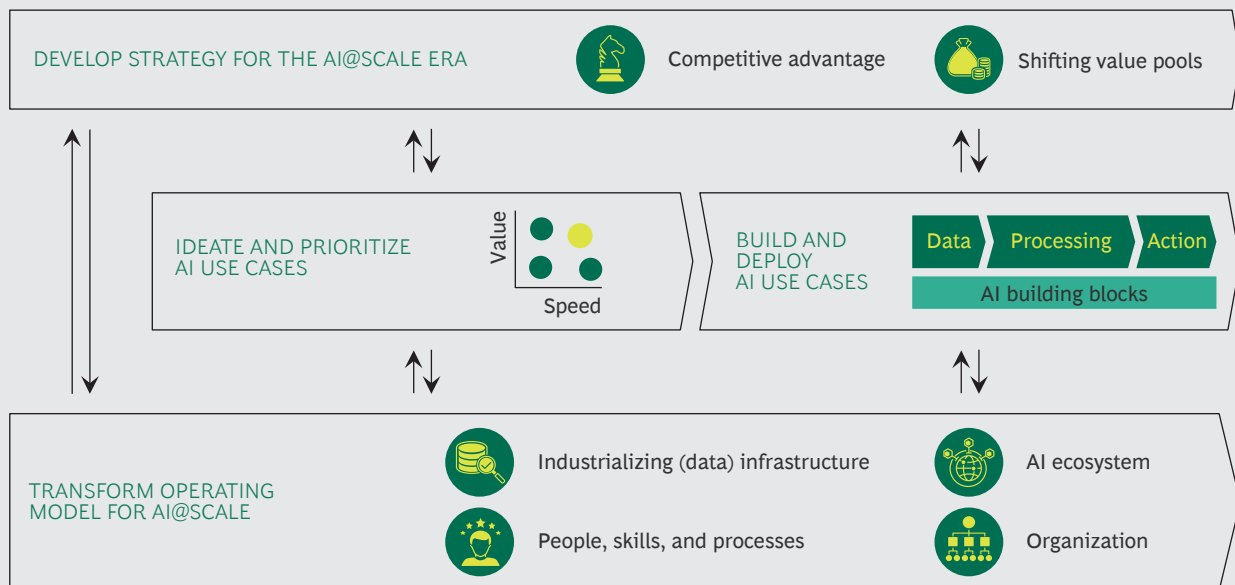
While strategy and AI use cases are not the focus of this publication, they are critical to AI@scale transformations. So that is where we start. (See Exhibit 1.)

STRATEGY FOR AI@SCALE

Strategy setting in an era of disruption is daunting, and AI is the epitome of disruption. So while it makes sense to initiate an AI strategy within or close to the current business model, companies must also take into account the evolution of value pools in the wider market as other players embrace AI.

The first step in that process involves studying the potential and overall effect of AI use cases for both internal processes and external offerings. When we ran such an analysis for the top ten banks worldwide, it generated the estimate (referenced earlier) of up to \$220 billion in additional annual operating earnings with AI. (See the sidebar “How AI Could Fatten the Bottom Line.”)

EXHIBIT 1 | Deploying AI@scale Requires Transformation



Source: BCG analysis.

The second step—understanding the evolution of value pools—is equally important. The rest of the world is not standing still, passively watching these rewards being realized. Other companies both within and outside a company’s traditional industry are focusing on how AI can help them attack value pools.

In a prior publication, we illustrated potential value shifts in the health care industry among biopharma companies, insurers, providers, medtech companies, new entrants from the technology industry, and consumers. (See *Putting Artificial Intelligence to Work*, BCG report, September 2017.) A foreshadowing of such shifts occurred in January 2018 when the stocks of major retail pharmacy chains and insurers dropped by up to 5% relative to the S&P 500 after Amazon, JPMorgan Chase, and Berkshire Hathaway announced their entry into health care. Financial services, media, retail, as well as mobility and logistics are other industries susceptible to major AI-triggered disruption.

HOW AI COULD FATTEN THE BOTTOM LINE

AI can drive tremendous value creation. We illustrate its potential by estimating the bottom-line impact it has had on the ten largest global banks through increased operating revenues and reduced operating expenses. AI can also improve risk management, but its value depends on country-specific regulations. In industries other than banking, moreover, AI can help optimize capital expenditures. The exhibit below shows the revenue gains and cost savings that AI can produce within corporate and retail banking.

AI can improve revenues by attracting new customers, increasing their share of wallet, and optimizing pricing. Some of the AI applications to boost revenues include microtargeting, personalized offerings, chatbots and robo-advisors, portfolio-based risk alerts, loan loss estimation, willingness-to-pay evaluations, and advanced fee schemes implementation. These measures could lift revenues (called operating income in banking) by \$120 billion to \$180 billion annually at the

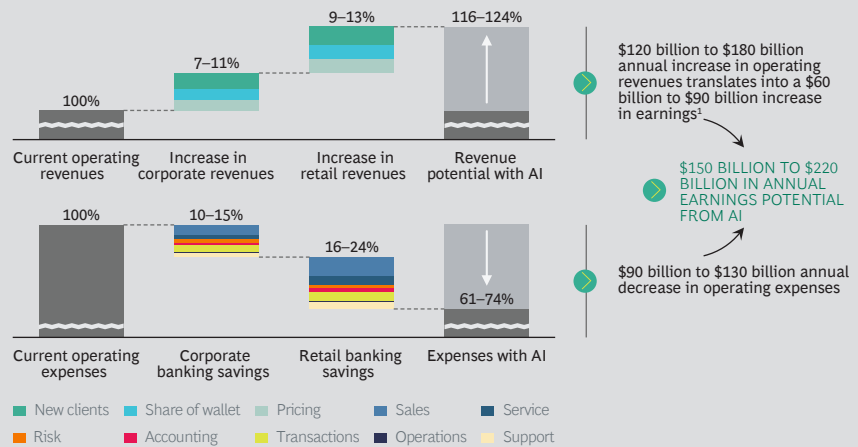
ten largest banks. Assuming a 50% ratio of costs to revenues (cost-to-income ratio in banking jargon) at the top ten banks, \$60 billion to \$90 billion would drop to the bottom line on an operating basis.

AI can lower costs by helping to automate a variety of front- and back-office activities including sales and service, risk management, account creation, transaction processing, operations, and support functions. Some common uses of AI to reduce costs include applications to address money laundering and fraud detection, workforce management, and expense auditing. AI-enabled savings could reach \$90 billion to \$130 billion at the ten largest banks.

The combination of revenue boosts and cost savings could generate an ongoing annual increase in operating earnings of \$150 billion to \$220 billion. Of course, banking is an industry in flux, so projections based on the current market environment may not be fully realized.

HOW AI COULD FATTEN THE BOTTOM LINE (continued)

AI Could Generate \$220 Billion in Additional Annual Earnings for the Top Ten Banks



Sources: BCG corporate banking digital excellence benchmark; SNL Financial; BCG analysis.

¹Earnings calculation assumes a 50% cost-to-revenue ratio for the top ten banks (referred to as the cost-to-income ratio in banking), not including additional potential from AI-driven operating-expense reductions for additional revenue.

Companies formulating their strategy need to consider both the potential of AI uses and the risk that today's value pools may shift.

Two warnings are in order in setting strategy.

It's not all about amassing data. While critical for the success of AI, internal data is not the Holy Grail. For one thing, companies increasingly will need to negotiate access to data outside their walls—from suppliers, partners, and customers, for example. Also, new approaches are emerging that avoid the need for massive troves of data. Generative adversarial networks (GANs), for instance, are virtual learning environments in which algorithms train one another to solve problems—such as creating and discriminating false pictures of faces. “Digital twins,” meanwhile, are virtual replicas of physical assets that enable algorithms to learn to optimize end-to-end operations, for example, without the same real-world data requirements of traditional AI.

It's also not enough to be agile and fast. Flexibility and speed are important, but they are not excuses to engage in a random walk through the AI wilds. To achieve competitive advantage, companies need to point in a specific direction. As research we have conducted with the *MIT Sloan Management Review* shows, the gap between companies believing in the transformational potential of AI and those actually having an AI strategy in place is stunning.

IDENTIFYING, PRIORITIZING, AND DEPLOYING AI USE CASES

Companies that have successfully piloted several AI use cases can get seduced into thinking that the move toward AI@scale is not that hard. Use cases can be powerful levers to boost performance. And they can help companies adopt agile test-and-learn methods; build transparency and cybersecurity measures into AI systems; and satisfy internal and regulatory requirements.

But isolated use cases provide limited guidance for the challenges of navigating through the AI paradox. They can sputter and grind to a halt when interacting at scale unless companies transform their operating models.

As a consequence, companies should initially prune the number of use cases, focusing on the most promising. (See the sidebar “How to Address AI Use Cases.”) There is peril—not safety—in numbers when scaling up AI use cases and pilots.

Preparing for AI@scale

Against that backdrop of strategy, priority setting, and piloting, let’s explore how companies can plan for a successful AI@scale transformation.

CREATING A ROBUST AI ARCHITECTURE

IT is often treated as a standalone function that is brought in to implement decisions already made by business leaders. This approach does not work with AI. If executives want to deploy machine intelligence to solve essential business problems, they must prioritize machine architecture equally to “people architecture” or traditional organizational issues. Otherwise, the entanglement concerns described earlier will disrupt their plans.

As companies move beyond the pilot phase, AI algorithms systemically interact with one another by learning from data, as explained earlier. Researchers at Google raised the entanglement issue several years ago, leading many prominent AI users to develop comprehensive machine-learning architectures. For example, Uber’s Michelangelo, in the words of the company, is “an end-to-end system that enables users across the company to easily build and operate machine learning systems at scale.” (See Exhibit 2.) Likewise, a leading European online fashion retailer has built an AI platform that allows individual AI systems to work together and reuse components.

Vendors such as Amazon, Google, IBM, and Microsoft have all now launched or announced platform offerings to AI users. Whether organizations buy one of these emerging platforms from a vendor or build their own, companies that want to deploy AI at scale need a rigorous and consistent framework and system to manage, document, and monitor workflow from data input to final action. In addition to addressing issues such as entanglement and reusability, for example, companies need to monitor the security of these systems as well as the integrity and compliance of automated actions generated by AI.

Moreover, companies need to ensure that they have the storage, computing, and bandwidth to handle multiple AI engines and their timely actions. Owing to its flexibility, the cloud is often a preferred option to address these needs. In industrial

Companies need a rigorous platform to manage workflow from data input to final action.

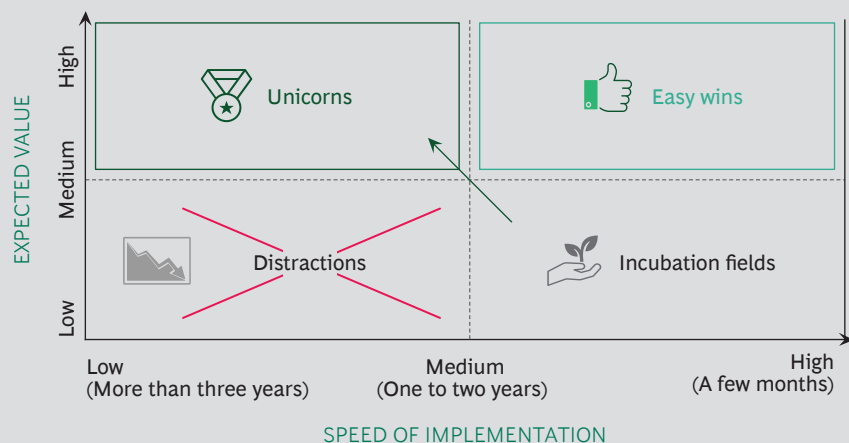
HOW TO ADDRESS AI USE CASES

Many companies are not yet applying a proven methodology for identifying and prioritizing use cases. Instead, they run many uncoordinated pilots and tests, then face tremendous challenges in overcoming the AI paradox, and often must revisit and restructure their previous work.

A global automaker addressed the problem upfront by carefully reviewing and evaluating a wide variety of options and then prioritizing the most critical functions and use cases for achieving the most important benefits. The company took a two-stage approach: it aims to hit an intermediate set of milestones by 2020 and reach full potential by 2025.

The exhibit below shows how to categorize use cases along the dimensions of speed of implementation and potential value creation. Easy wins should be a priority because they are quick and high value but also rare. The unicorn category promises the most potential but takes the longest time to implement. Smaller initiatives that could eventually become unicorns (what we call “incubation fields”) make sense as ways to gain knowledge and experience and embark on the journey. All other use cases are distractions and should be avoided or discontinued as soon as possible.

By Prioritizing Use Cases, Companies Can Focus on Easy Wins and Unicorns



Source: BCG analysis.

EXHIBIT 2 | Uber's Michelangelo Platform Helps Harmonize AI Deployment

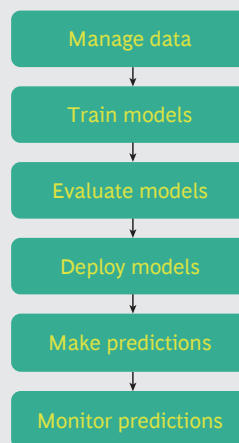
THE RISKS OF SCATTERED AND ISOLATED AI DEPLOYMENT

- ! No central data and pipeline management for AI systems
- ! Fragmented storage of features, models, and outputs
- ! Few employees with required technical know-how
- ! Unsystematic deployment process for AI systems

HOW UBER'S MICHELANGELO INDUSTRIALIZES AI

- Central data lake with data source and pipeline management
- AI platform that handles various use cases with central feature store
- Many teams building and deploying AI models at once
- End-to-end centralized deployment support

MICHELANGELO'S CRITICAL END-TO-END MANAGEMENT OF AI ENVIRONMENT



Sources: "Hidden Technical Debt in Machine Learning Systems," *Proceedings of the 28th International Conference on Neural Information Processing Systems*, Sculley D, et al., 2015; Uber website; BCG analysis.

automation and Internet of Things networks, however, latency and bandwidth constraints may prevent the cloud from serving as a complete solution. These settings require novel structures, such as edge computing, in which part of the processing power is kept closer to the action in the periphery.

Ultimately, the intelligent algorithms are a relatively small part at the core of the overall system. Consequently, many companies expanding their AI operations need only a small number of data scientists and AI experts at the core but require a large number of data and systems engineers with AI experience to ensure the performance and resilience of the pipeline and peripheral systems. For instance, with experienced data engineers and architects, a leading oil and gas company has built its own integrated-operations center, leveraging AI to enable predictive maintenance and optimize production. The projected efficiency gains amount to several billion dollars over the next few years.

STRUCTURING AN ECOSYSTEM

The question is not if companies will work with vendors and partners but how. At least for the short term, it is their only way to access necessary AI talent and capabilities. Even companies with strong AI backgrounds rely to some extent on vendors.

With AI, however, these relationships could jeopardize core sources of competitive advantage for the companies. Vendors need to train their AI tools, often using sensitive client data.

Companies can work with AI vendors in many ways, from outsourcing an entire process to buying selected services, seeking help in building in-house solutions, or training

internal staff. Executives should view these options in light of two questions:

- How valuable is the process or offering to our future success?
- How strong is our ownership, control, or access to high-quality, unique data, relative to the AI vendor's?

By analyzing the AI landscape this way, companies will discover that their AI efforts land in one of four quadrants that each requires a different strategy:

- **Commodities (low-value potential, low differentiated data access).** Manage vendor relationships to reduce costs and improve performance in support functions and elsewhere.
- **Danger Zones (high-value potential, low differentiated data access).** Develop or acquire competitive data sources to turn danger zones into gold mines.
- **Hidden Opportunities (low-value potential, high differentiated data access).** Leverage technical expertise of vendors to generate quick wins and insights.
- **Gold Mines (high-value potential, high differentiated data access).** Capitalize on superior data access in strategically relevant areas and develop AI independently.

Beyond the critical issues of trust and confidentiality, the overall commercial offerings of vendors—including their newly launched platforms—differ widely in capabilities, integration abilities, and ease of use. So companies need sufficiently solid AI capabilities to select and manage these relationships. Many business leaders ultimately choose to hire an independent advisor to assess the tradeoffs while they build up the internal capability to bring work in-house.

DEVELOPING PEOPLE, SKILLS, AND PROCESSES

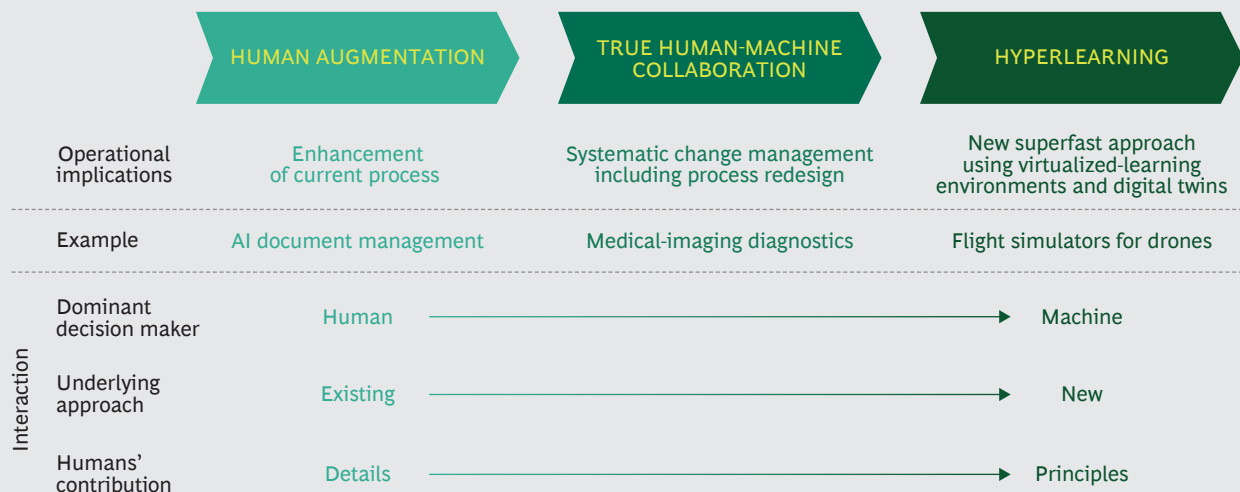
Companies moving toward AI@scale should carefully consider how the transformation will affect their workforce through the creation of skills shortages, elimination of jobs, or both. This complex topic requires intensive strategic workforce planning, retraining, and process re-engineering to accommodate a human-machine world.

The most immediate need is critical technical skills. While many companies focus on hiring data scientists, they are encountering a greater shortage of people who have both business skills and an understanding of AI, as well as systems and data engineers, as mentioned earlier.

Longer term, companies need to understand the fundamental and nuanced shifts in the workplace that occur as humans and machines work side by side. (See Exhibit 3.) When AI is focused on the augmentation of employees in their current roles, changes to the workforce and processes are relatively modest. Indeed, the introduction of augmentation in many internal processes and service functions often improves employees' experience. In financial services, for example, the use of AI is reducing the burden while enhancing the quality of many processes.

Companies need to understand the shifts in the workplace that occur as humans and machines work side by side.

EXHIBIT 3 | The Patterns and Consequences of Human-Machine Interactions



Source: BCG analysis.

These changes become more extreme as processes start being reimagined for true human-machine collaboration. In radiology, doctors and hospitals are already starting to prepare their processes and business models for a time when regulators accept machines as a second opinion—and ultimately a first opinion—for diagnostics, with a human in the loop to correct potential errors.

Finally, in what we call “hyperlearning,” machines are capable of teaching themselves in a virtual environment with little or no human intervention and without relying on massive data input from the real world. As discussed earlier, GANs and digital twins are such environments. It is conceivable that algorithms will not need human training or even real-world data in these environments in the near future. This evolution has already occurred in the relatively restricted environment of chess but also, for instance, in flight simulators for drones and end-to-end supply chain optimization.

As companies move along this continuum, they will need to address new and novel change management and reskilling demands. Renault has already embarked on this path with its digital transformation and with the creation of its Digital Hub, a large center dedicated to the development of digital products, training, and innovation.

DESIGNING GOVERNANCE AND ORGANIZATIONAL STRUCTURES

Governance and organization of any major transformation are heavily dependent on context and goals. But all AI@scale transformations require the general three-dimensional structure mentioned in the introduction. (See Exhibit 4.)

Corporate Center. Certain types of expertise and governance should be centralized. These include excellence hubs that house experts in many of the (horizontal) AI building blocks—such as machine vision, natural-language processing, and some general frontier topics in machine learning—and support AI initiatives and vendor assessment throughout the company.

EXHIBIT 4 | AI@scale Requires Changes in Organization and Governance

CORPORATE GOVERNANCE AND SUPPORT HUBS



AI center of
excellence



Data
governance team



IT architects

CENTRALIZED LEARNING AND USE-CASE BUILDING



Business owners
and use-case
teams



SWAT teams
for change

DECENTRALIZED ACTION AND IMPLEMENTATION



Organizations
affected by AI

Source: BCG analysis.

Data is the raw material of AI, but it also contains some of the company's most sensitive information. A world-class data governance function, which sets data permissions across the organization, is critical in an AI world, to ensure both competitive advantage and regulatory compliance.

Finally, data architects, who manage global topics such as industrialization frameworks and platforms as well as vendors, should be centralized along with cybersecurity experts.

Business Units and Functions. As a general rule, cross-functional agile teams situated within business units or functions should be responsible for the development of AI-enabled processes, products, and services. These teams should include both AI experts and topic specialists. Roving “SWAT teams,” including HR experts and change specialists, should also be available to help decentralized units affected by AI use cases with implementation, training, and other issues.

Field Level. The employees responsible for managing the processes and actions modified by AI are, of course, decentralized, located close to the market, shop floor, or field. These units will need to understand the new tools and how they affect processes and skill requirements and will carry the principal burden of change. The SWAT teams will be available to bring these people up to speed on how to leverage the new technology and assist with the transformation.

Pulling Together the Transformation Program

An AI@scale transformation should occur through a series of top-down and bottom-up actions to create alignment, buy-in, and follow-through. This has ensured the successful industrialization of AI across companies and their value chains:

- **AI Ambition and Maturity Assessment.** This top-down step establishes the overall context of the transformation and helps prevent the company from

pursuing disconnected AI pilots. The maturity assessment is typically based on a combination of surveys and interviews.

- **Evaluation of AI Initiatives and the Operating Model.** This bottom-up step provides a baseline of current AI initiatives. It should include goals, business cases, accountabilities, work streams, and milestones in addition to an analysis of data management, algorithms, performance metrics, and cybersecurity. A review of the current AI operating model should also be conducted at this stage.
- **Priority Setting and Gap Analysis.** The next top-down step prioritizes AI initiatives, focusing on easy wins and unicorns. This step also identifies the required changes to the operating model.
- **Outline of AI@scale Transformation Program.** This top-down step consists of both the transformation roadmap, including the order of initiatives to be rolled out, and the creation of a program management office to oversee the transformation.
- **Detailed Implementation Planning of AI@scale Program.** The last step covers implementation, detailing the work streams, responsibilities, targets, milestones, and resources.

By systematically moving through these steps, the implementation of the AI@scale transformation will proceed with much greater speed and certainty. Still, companies must be aware that the transformation can take one to three years, depending on the complexity of the overall organization.

AI IS A promising and alluring technology. Many companies have had early success with pilots but have fallen victim to the AI paradox. An AI@scale program helps ensure that companies will methodically address the multifaceted dimensions of a full transformation and reap the tremendous benefits.

For Further Reading

The authors and their Boston Consulting Group colleagues have published other reports and articles that may be of interest to senior financial executives. Recent examples include those listed here.

AI in the Factory of the Future

A Focus by The Boston Consulting Group, April 2018

Towards a Reskilling Revolution

A report by The Boston Consulting Group and the World Economic Forum, January 2018

Reshaping Business with Artificial Intelligence

An article by The Boston Consulting Group and the *MIT Sloan Management Review*, September 2017

Why AI Matters—No Magic Wand but Also No Simple Tool

An article by The Boston Consulting Group's Philipp Gerbert, *Business 20 Dialogue*, February 2017

Competing in the Age of Artificial Intelligence

An article by The Boston Consulting Group, January 2017

About the Authors

Philipp Gerbert is a senior partner and managing director in the Munich office of The Boston Consulting Group and a BCG Henderson Institute Fellow exploring the impact of AI on business. He coleads the firm's AI@scale initiative and leads its global digital strategy topic. You may contact him by email at gerbert.philipp@bcg.com.

Sukand Ramachandran is a partner and managing director in the firm's London office. He is a core group member of BCG's Technology Advantage practice and coleads the AI@scale initiative. You may contact him by email at ramachandran.sukand@bcg.com.

Jan-Hinnerk Mohr is a principal in BCG's Berlin office. He is a BCG Henderson Institute ambassador exploring the impact of AI on business and a core member of the firm's Technology, Media & Telecommunications practice. You may contact him by email at mohr.jan@bcg.com.

Michael Spira is a project leader in the firm's Munich office. He is also a BCG Henderson Institute ambassador exploring the impact of AI on business and a core member of BCG's Energy practice. You may contact him by email at spira.michael@bcg.com.

Acknowledgments

The authors thank Sebastian Steinhäuser, Henri Humpert, Johannes Knebel, and Lukas Rapp for support in compiling this report. They also thank Tilman Buchner, Ralf Dreischmeier, Sylvain Durranton, Christoph Gauger, Antoine Gourévitch, Andreas Gräf, Christoph Hartmann, Derek Kennedy, Rolf Kilian, Ryoji Kimura, Andreas Lundmark, Roger Premo, Arun Ravindran, Martin Reeves, Thomas Reichert, Michael Ruhl, and Daniel Schlecht for their invaluable comments and contributions, as well as Ralf Herbrich, director of machine learning at Amazon, for his remarks on an early draft.

The authors also are grateful to Mark Voorhees for writing assistance. In addition, they thank Katherine Andrews, Gary Callahan, Siobhan Donovan, Kim Friedman, Abby Garland, Sean Hourihan, and Shannon Nardi for editorial and production support.

To find the latest BCG content and register to receive e-alerts on this topic or others, please visit bcg.com.

Follow The Boston Consulting Group on Facebook and Twitter.

© The Boston Consulting Group, Inc. 2018. All rights reserved.
6/18

