Harvard Business Review Special Issue

The Best of HBR
Winter 2021 HBR.org

How AI Is Changing Work

What you need to know about automation, machine learning, and the future of jobs

Publicis EtNous

HERMÈS
PARIS
Objects connect.

If you're customer-obsessed.

We should talk.

Talkdesk is the only cloud contact center solution built for a better customer experience.

Talkdesk® is a global customer experience leader for customer-obsessed companies. Talkdesk CX Cloud™ is an end-to-end customer experience solution that uses pragmatic AI to deliver a better way for businesses and customers to engage with one another, with customer self-service, agent empowerment, fraud prevention, and enhanced automation. Over 1,800 innovative companies around the world, partner with Talkdesk to deliver a better way to great customer experience. Learn more and request a demo at www.talkdesk.com.

talkdesk®

Experience. A better way.

talkdesk.com

Managing Our New Robot Overlords

DESPITE THE GREAT promise that AI holds for companies looking to create better, faster experiences for customers, adopting these technologies is one of the most complex management tasks of a generation. We've assembled this special issue to help leaders take control of the process and—maybe—the technology.

From an automated Amazon that predicts purchases to bracelets that keep a finance firm's traders from making purchase decisions on the basis of heightened emotions, applying AI technologies is a critical competitive advantage. The first section of this issue describes these and other company stories and the new ways of thinking about strategy that they're ushering in.

Even once you have a sense of how you could serve your market with AI, adopting and scaling these technologies within your organization is a behemoth of an undertaking. The conventional wisdom—to start small—has given way to a recognition that organizations need a fundamental shift in their approach to data to do even that well. Turn to the second section of this issue for more on how to onboard AI.

Especially as the pandemic has accelerated automation, though, the big question on everyone's minds is whether all that onboarding eventually means that the bots are going to take all our jobs. The promising answer from most experts is that there will certainly be roles for humans even in a machine-powered future; the real question is what those jobs will be, how companies can prepare their workforces to do them, and how humans and AI will work together. The next two sections address the future of jobs.

Finally, the question at the crux of any discussion about AI is this: Are we managing the machines, or are they managing us? Leaders need to set boundaries for the bots, ethical and otherwise. The last section of this issue addresses the ways that AI's tentacles can reach dangerously beyond what we expect—so we can learn how to keep them reined in.

– The Editors

I DISCOVERED

how to develop and deliver life-saving drugs to people faster

It's not your data. It's how you use it. Whether creating shared data experiences to innovate rapidly and deliver life-saving drugs to people faster or developing energy efficient projects that lead to reduced carbon emissions, transforming your business requires the power of artificial intelligence and machine learning to translate complex data sets into clear and actionable insights. Cloudera's hybrid data platform accelerates data analytics at every stage of the data lifecycle, with security and governance built in, to move your business forward.

Learn more at **cloudera.com/datamovesyou**

#datamovesyou

CLOUDERA
Data That Moves You

WINTER 2021

Contents

Culled by the editors of *Harvard Business Review* from the magazine's rich archives, these articles are written by some of the world's leading management scholars and practitioners. To help busy leaders apply the concepts, they are accompanied by "Idea in Brief" summaries.

Competing with AI

Competing in the Age of AI | 12
Marco Iansiti and Karim R. Lakhani

The Business of Artificial Intelligence | 20
Erik Brynjolfsson and Andrew McAfee

How AI Will Change Strategy: A Thought Experiment | 30
Ajay Agrawal, Joshua Gans, and Avi Goldfarb

Why Companies That Wait to Adopt AI May Never Catch Up | 32
Vikram Mahidhar and Thomas H. Davenport

Why AI That Teaches Itself to Achieve a Goal Is the Next Big Thing | 35
Kathryn Hume and Matthew E. Taylor

Three Ways AI Is Getting More Emotional | 39
Sophie Kleber

How Hospitals Are Using AI to Battle Covid-19 | 41
Kelley A. Wittbold, Colleen Carroll, Marco Iansiti, Haipeng (Mark) Zhang, and Adam B. Landman

Adopting AI

Artificial Intelligence for the Real World | 44
Thomas H. Davenport and Rajeev Ronanki

Getting AI to Scale | 54
Tim Fountaine, Brian McCarthy, and Tamim Saleh

MICHELLE D'URBANO

Cover illustration by **NISHANT CHOKSI**
Photograph of robot arm by AlexLMX/Getty Images

Today, almost everything is a project, and everyone is a project manager.

Padman Ramankutty
MANAGING DIRECTOR, ACCENTURE

Today, everyone needs to drive work forward, adapt to change, work across silos, and build the perfect solution to get work done. That's the new dynamic of project management. And it all happens on Smartsheet, the platform for dynamic work.

Get a free book on dynamic project management or start doing it at **smartsheet.com/bookoffer**

Our Commitment to Sustainability
We're proud that the paper we use in our print magazine is certified under the Sustainable Forestry Initiative® program, meaning that it comes from responsibly managed sources and is a renewable resource.

SUSTAINABLE FORESTRY INITIATIVE
Certified Sourcing
www.sfiprogram.org
SFI-00993

Inside Facebook's AI Workshop | 62
Scott Berinato

A Better Way to Onboard AI | 70
Boris Babic, Daniel L. Chen, Theodoros Evgeniou, and Anne-Laure Fayard

How to Choose Your First AI Project | 78
Andrew Ng

Is Your Data Infrastructure Ready for AI? | 80
Seth Earley and Josh Bernoff

The Future of Jobs

Collaborative Intelligence: Humans and AI Are Joining Forces | 84
H. James Wilson and Paul R. Daugherty

How Will AI Change Work? Here Are Five Schools of Thought | 94
Mark Knickrehm

AI Should Augment Human Intelligence, Not Replace It | 97
David De Cremer and Garry Kasparov

Why Robots Won't Steal Your Job | 100
Nahia Orduña

How Amazon Automated Work and Put Its People to Better Use | 104
Alex Kantrowitz

What We Often Get Wrong About Automation | 106
Ravin Jesuthasan and John Boudreau

How to Harness the Digital Transformation of the Covid Era | 109
Federica Saliola and Asif M. Islam

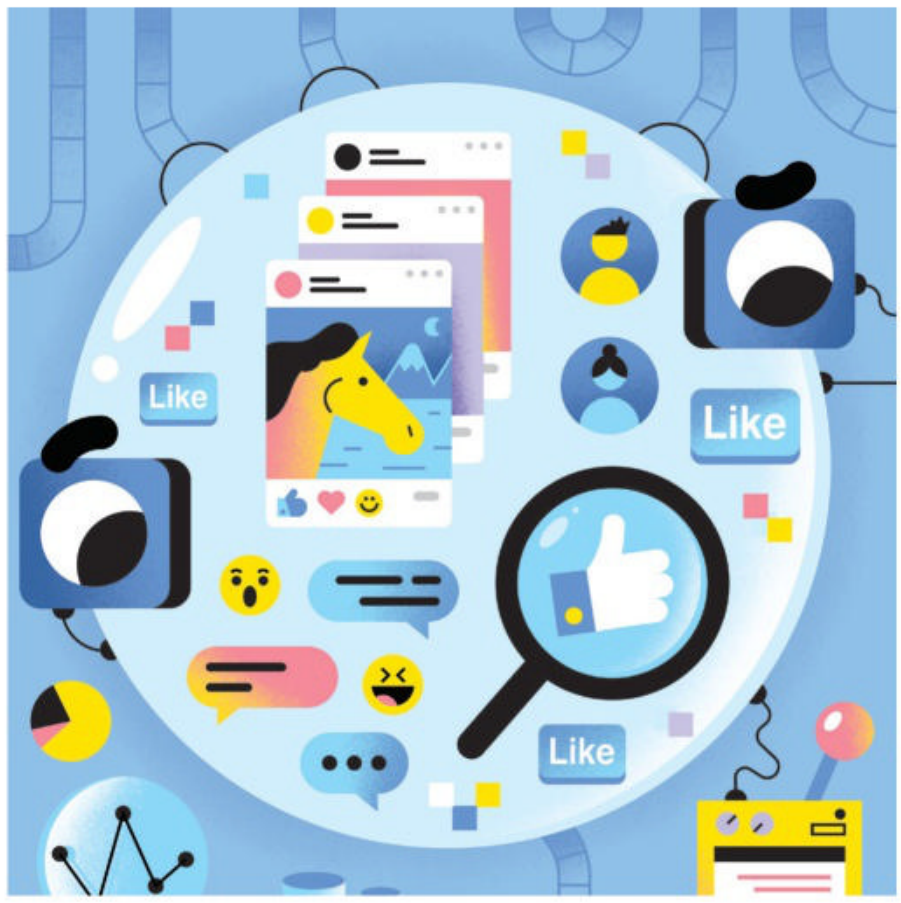
Working with Machines

Learning to Work with Intelligent Machines | 112
Matt Beane

Algorithms Need Managers, Too | 120
Michael Luca, Jon Kleinberg, and Sendhil Mullainathan

The Future of AI Will Be About Less Data, Not More | 126
H. James Wilson, Paul R. Daugherty, and Chase Davenport

The Secret to AI Is People | 128
Nada R. Sanders and John D. Wood



Keeping AI from Going Wrong

When Machine Learning Goes Off the Rails | 132
Boris Babic, I. Glenn Cohen, Theodoros Evgeniou, and Sara Gerke

What Will Happen When Your Company's Algorithms Go Wrong? | 140
Roman V. Yampolskiy

What People Hate About Being Managed by Algorithms, According to a Study of Uber Drivers | 142
Mareike Möhlmann and Ola Henfridsson

Why Business Leaders Need to Understand Their Algorithms | 145
Mike Walsh

If Your Company Uses AI, It Needs an Institutional Review Board | 147
Reid Blackman

Executive Summaries | 152

ENISAURUS

Let's make the workplace worth smiling about.

HR solutions that will keep your people happy.

UKG

Our purpose is people

Learn more at UKG.com/whyukg

Harvard Business Review

AI is here. Are you ready?

How will AI impact your industry, company, and career? Get up to speed on the latest research and intel with HBR's new **Artificial Intelligence Set: How to Stay Competitive in an AI World**.

In this specially priced collection, you'll explore how to:

- Lay the groundwork for transformation
- Incorporate AI into your business strategy
- Reimagine work processes
- Manage people in this brave new world
- Leverage AI to be a market leader

Artificial Intelligence Set: How to Stay Competitive in an AI World
PRODUCT #1121BN · 5 ITEMS · $125 · DIGITAL FORMAT

A savings of more than 25% off the individual components!

THIS DIGITAL SPECIAL COLLECTION INCLUDES 5 EBOOKS + TOOLS

AVAILABLE EXCLUSIVELY AT:
store.hbr.org
1-800-988-0886 OR +1-617-783-7500

Harvard Business Review

EDITOR IN CHIEF Adi Ignatius

EDITOR, HBR Amy Bernstein

EDITOR, HBR.ORG Maureen Hoch

EDITORIAL DIRECTOR, HBR PRESS Melinda Merino

CREATIVE DIRECTOR John Korpics

EXECUTIVE EDITORS Alison Beard, Daniel McGinn, Sarah Moughty, Ania G. Wieckowski

EDITORIAL

SENIOR EDITORS Laura Amico, Scott Berinato, David Champion *Paris,* Paige Cohen, Kevin Evers, Gretchen Gavett, Eben Harrell, Jeff Kehoe, Scott LaPierre, Toby Lester, Juan Martinez, Amy Meeker, Curt Nickisch, Steven Prokesch, Vasundhara Sawhney, Thomas Stackpole, Erica Truxler

MANAGING EDITOR, HBR PRESS Allison Peter

SENIOR ASSOCIATE EDITORS Courtney Cashman, Susan Francis, Dave Lievens

ASSOCIATE EDITORS Dagny Dukach, Stefanie Fernández, Ramsey Khabbaz, Emma Waldman

ARTICLES EDITORS Christina Bortz, Susan Donovan, Martha Lee Spaulding

EDITORIAL AUDIENCE DIRECTOR Nicole D. Smith

AUDIENCE ENGAGEMENT EDITOR Kelsey Gripenstraw

SENIOR AUDIO PRODUCERS Amanda Kersey, Anne Noyes Saini

MULTIMEDIA PRODUCER Andy Robinson

ASSISTANT EDITORS Rakshitha Ravishankar, Alicyn Zall

EDITORIAL COORDINATOR Shanze Hasan

STAFF ASSISTANT Christine C. Jack

CONTRIBUTING EDITORS Sarah Cliffe, Karen Dillon, Amy Gallo, Jane Heifetz, Paul Hemp, John Landry, Gardiner Morse, Andrew O'Connell, Anand P. Raman, Dana Rousmaniere

CONTRIBUTING STAFF Leslie Cauldwell, Kathryn K. Dahl, Sarabeth Fields, Kelly Messier, Annie Noonan, JM Olejarz, Debbie White

DESIGN

DESIGN DIRECTORS Stephani Finks *HBR Press,* Susannah Haesche *HBR,* Marta Kusztra *Product & UX*

ASSOCIATE DESIGN DIRECTOR Karen Player *Multimedia*

SENIOR GRAPHICS EDITOR Macaulay Campbell

SENIOR DESIGNER Emily Caulfield

SENIOR PHOTO EDITOR Sasha Patkin

DESIGNERS Josephine Massey *UX,* Jeff Paul *Product*

CONTRIBUTING DESIGNERS Yuliya Bas, Kristen Nathan

PRODUCTION

EDITORIAL PRODUCTION DIRECTOR Dana Lissy

SENIOR PRODUCTION EDITORS Jennifer Waring, Christine Wilder

PRODUCTION EDITORS Victoria Desmond, Jodi Fisher, Anne Starr

SENIOR PRODUCTION SPECIALIST Robert Eckhardt

PRODUCTION SPECIALIST Alexie Rodriguez

CONTRIBUTING STAFF Kristin Murphy Romano

EDITORIAL ADVISORY BOARD

Bharat Anand, Azeem Azhar, John Battelle, Marcus Buckingham, Nicco Mele, Vivek Shah

EDITORIAL OFFICES
60 Harvard Way, Boston, MA 02163
617-783-7410 | fax 617-783-7493
HBR.org

WINTER 2021
HARVARD BUSINESS REVIEW SPECIAL ISSUE
ISSN 2689-5447
Printed in the U.S.A.

Copyright 2021 Harvard Business School Publishing Corporation. All rights reserved.

GROUP PUBLISHER Sarah McConville

VICE PRESIDENT, GLOBAL ADVERTISING SALES
Gail Day

MANAGING DIRECTOR, DIGITAL PRODUCT STRATEGY
Jim Bodor

VICE PRESIDENT, NATIONAL ADVERTISING & MARKETING
Craig Catalano

SENIOR DIRECTOR, TECHNOLOGY
Kevin Newman

ADVERTISING
MANAGING DIRECTOR, ANALYTIC SERVICES Alex Clemente
DIRECTOR, INTERNATIONAL ADVERTISING & PARTNERSHIPS Daniel Cohen
SENIOR EDITOR, RESEARCH & SPECIAL PROJECTS Todd Pruzan
MANAGING EDITOR, ANALYTIC SERVICES Anthony Baldo
SENIOR RESEARCHER, ANALYTIC SERVICES Cynthia Talkov Boyd
WRITER/EDITOR, ANALYTIC SERVICES Danielle Warren
DIRECTOR, MARKETING Yasir Salem
ASSOCIATE DIRECTOR, MARKETING Samantha Barry
MARKETING MANAGER Mary Callaghan
ASSISTANT MANAGER, MARKETING Alexandra Shore
MARKETING COORDINATOR Tendayi Peyton
CONTRIBUTING CONSULTANT MaryAlice Holmes

MARKETING & COMMUNICATIONS
DIRECTOR, CONSUMER MARKETING Nini Diana
DIRECTOR, COMMUNICATIONS Amy Poftak
ASSISTANT DIRECTOR, E-COMMERCE Carol Concannon
SENIOR MANAGER, CONSUMER MARKETING Corrine Callahan
SENIOR MANAGER, ACQUISITION Josh Getman
SENIOR MANAGER, CUSTOMER RETENTION & LOYALTY Caty Jackson

PRODUCT MANAGEMENT
DIRECTOR, PRODUCT MANAGEMENT Anne Bartholomew
SENIOR PRODUCT MANAGER, AUDIO & VIDEO Adam Buchholz
SENIOR AGILE PROJECT MANAGER Don Doucette
SENIOR TECHNICAL PRODUCT MANAGER Frederic Lalande
PRODUCT MANAGERS Jon Goldberg, Dhanan Lathia

BUSINESS & CUSTOMER ANALYTICS
SENIOR DIRECTOR, BUSINESS ANALYTICS & INSIGHTS Jessica Avery
DIRECTOR, CUSTOMER ANALYTICS & INSIGHTS Shelly McArdle
SENIOR CUSTOMER & MARKETING ANALYST Ellen Bonaccorsi
SENIOR RESEARCHER, EXPERIMENTATION & ANALYTICS Abigail Dawson
RESEARCHERS, EXPERIMENTATION & ANALYTICS Chris Chaky, Amanda Ludden
SENIOR EMAIL MARKETING PRODUCTION SPECIALIST Frances Lee
EMAIL MARKETING COORDINATOR Milo Goodman

FINANCE & OPERATIONS
SENIOR DIRECTOR, FINANCE & OPERATIONS Greg St. Pierre
SENIOR DIRECTOR, INTERNATIONAL STRATEGY Kate Griffin
ASSOCIATE DIRECTOR, OPERATIONS & CUSTOMER EXPERIENCE Greg Daly
CUSTOMER EXPERIENCE MANAGER Danielle Weber
MANAGER, FINANCIAL PLANNING & ANALYSIS Krishna Dudhwewala
FINANCIAL ANALYST Joel Trudel
DATA INTEGRITY & OPERATIONS SPECIALIST Edward D. Domina IV

HBR PRESS
COMMERCIAL DIRECTOR, ASSOCIATE PUBLISHER Erika Heilman
DIRECTOR, MARKETING & COMMUNICATIONS Julie Devoll
DIRECTOR, LICENSING & INTERNATIONAL SALES Jon Shipley
ASSOCIATE DIRECTOR, MEDIA & COMMUNICATIONS Sally Ashworth
ASSOCIATE DIRECTOR, PUBLICITY Felicia Sinusas
MANUFACTURING MANAGER Rick Emanuel
SENIOR SALES OPERATIONS MANAGER Brian Galvin
BUSINESS DEVELOPMENT MANAGER, DIRECT SALES Lindsey Dietrich
MARKETING AND COMMUNICATIONS SPECIALIST Alexandra Kephart

TECHNOLOGY
TECHNICAL ARCHITECT Stepan Sokolov
SENIOR WEB DEVELOPER Matt Wagner
SENIOR RELEASE ENGINEER Matthew Han
SENIOR APPLICATION DEVELOPER Rollin Crittendon
FRONT-END ENGINEER Caitlin Smith

A NOTE TO READERS
The views expressed in articles are the authors' and not necessarily those of *Harvard Business Review*, Harvard Business School, or Harvard University. Authors may have consulting or other business relationships with the companies they discuss.

LIBRARY ACCESS
Libraries offer online access to current and back issues of *Harvard Business Review* through EBSCO host databases.

ARTICLE REPRINTS
To purchase reprints of *Harvard Business Review* articles, go to HBR.org.

WORLDWIDE ADVERTISING OFFICES

NEW YORK
3 Columbus Circle, Suite 2210
New York, NY 10019
212-872-9280 | fax 212-956-0933
Maria A. Beacom *Senior Account Manager; Director, North American Research*
Michael Valecce *New England & NY Luxury Account Manager*

CHICAGO
847-466-1525 | fax 847-466-1101
James A. Mack
Central U.S. Sales Director

MIDWEST & SOUTHEAST
312-867-3862 | cell 312-401-2277
Samuel K. White
Midwest & Southeast Sales Director

LONDON
44-0-7890-608-707
Alexander Bart
Associate Director, Ad Sales
44-0-7538-236-722
Jack Elia *Sales Account Manager*

LOS ANGELES 310-546-3757
SAN FRANCISCO 415-986-7762
BEIJING 86-10-6551-0663
FRANCE 33-01-4643-0066
HONG KONG 852-237-52311
INDIA 212-872-9291
JAPAN 81-03-3541-4166
KOREA 82-2-730-8004
SOUTH AMERICA 55-11-98160-0177
UAE 971-4-228-7708
UNITED KINGDOM 44-20-7291-9129

For all other inquiries, please call 212-872-9280 or visit HBR.org/hbr-advertising-sales.

SUBSCRIPTION SERVICES

UNITED STATES & CANADA
800-274-3214
Harvard Business Review
P.O. Box 37457; Boone, IA 50037-0457
HBR.org/subscriberservices

ALL OTHER COUNTRIES
Asia Pacific region: 61-2-9158-6127
All other regions: 44-1858-438412
Harvard Business Review
Tower House, Lathkill Street
Market Harborough LE16 9EF
United Kingdom
www.subscription.co.uk/hbr/help

RATES PER YEAR
United States $119 | Canada US$139
International US$165 | Mexico US$139

SUBMISSIONS
We encourage prospective authors to follow HBR's "Guidelines for Authors" before submitting manuscripts.

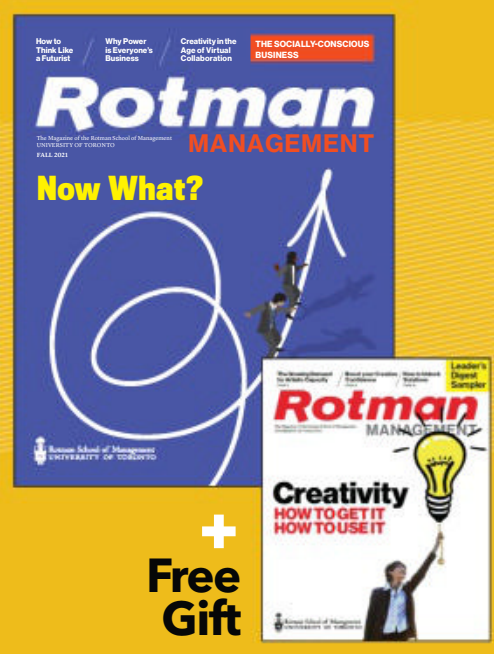
To obtain a copy, please visit HBR.org; write to: The Editor, *Harvard Business Review*, 60 Harvard Way, Boston, MA 02163; or email hbr_editorial@HBR.org.

Unsolicited manuscripts will be returned only if accompanied by a self-addressed stamped envelope.

Rotman MANAGEMENT

"Read cover to cover. Superb."

Tom Peters
Author, *In Search of Excellence*; Thinkers50 Hall of Fame



+ **Free Gift**

Filled with thought-provoking ideas and insights from leading global academics and management practitioners. Published by Canada's leading business school.

Subscribe today and receive our mini issue on creativity FREE!

Just $49.95 CAD/year
Available in print or digital

www.rotman.utoronto.ca/hbr

Rotman School of Management
UNIVERSITY OF TORONTO

ORIGINALLY PUBLISHED JANUARY–FEBRUARY 2020

Competing in the Age of AI

How machine intelligence changes the rules of business

→ by MARCO IANSITI and KARIM R. LAKHANI

IN 2019, JUST five years after the Ant Financial Services Group was launched, the number of consumers using its services passed the one billion mark. Spun out of Alibaba, Ant Financial uses artificial intelligence and data from Alipay—its core mobile-payments platform—to run an extraordinary variety of businesses, including consumer lending, money market funds, wealth management, health insurance, credit-rating services, and even an online game that encourages people to reduce their carbon footprint. The company serves more than 10 times as many customers as the largest U.S. banks—with less than one-tenth the number of employees. At its last round of funding, in 2018, it had a valuation of $150 billion—almost half that of JPMorgan Chase, the world's most valuable financial-services company.

Illustration by **MICHELLE D'URBANO**

Unlike traditional banks, investment institutions, and insurance companies, Ant Financial is built on a digital core. There are no workers in its "critical path" of operating activities. AI runs the show. There is no manager approving loans, no employee providing financial advice, no representative authorizing consumer medical expenses. And without the operating constraints that limit traditional firms, Ant Financial can compete in unprecedented ways and achieve unbridled growth and impact across a variety of industries.

The age of AI is being ushered in by the emergence of this new kind of firm. Ant Financial's cohort includes giants like Google, Facebook, Alibaba, and Tencent, and many smaller, rapidly growing firms, from Zebra Medical Vision and Wayfair to Indigo Ag and Ocado. Every time we use a service from one of those companies, the same remarkable thing happens: Rather than relying on traditional business processes operated by workers, managers, process engineers, supervisors, or customer service representatives, the value we get is served up by algorithms. Microsoft's CEO, Satya Nadella, refers to AI as the new "runtime" of the firm. True, managers and engineers design the AI and the software that makes the algorithms work, but after that, the system delivers value on its own, through digital automation or by leveraging an ecosystem of providers outside the firm. AI sets the prices on Amazon, recommends songs on Spotify, matches buyers and sellers on Indigo's marketplace, and qualifies borrowers for an Ant Financial loan.

The elimination of traditional constraints transforms the rules of competition. As digital networks and algorithms are woven into the fabric of firms, industries begin to function differently and the lines between them blur. The changes extend well beyond born-digital firms, as more-traditional organizations, confronted by new rivals, move toward AI-based models too. Walmart, Fidelity, Honeywell, and Comcast are now tapping extensively into data, algorithms, and digital networks to compete convincingly in this new era. Whether you're leading a digital start-up or working to revamp a traditional enterprise, it's essential to understand the revolutionary impact AI has on operations, strategy, and competition.

The AI Factory

At the core of the new firm is a decision factory—what we call the "AI factory." Its software runs the millions of daily ad auctions at Google and Baidu. Its algorithms decide which cars offer rides on Didi, Grab, Lyft, and Uber. It sets the prices of headphones and polo shirts on Amazon and runs the robots that clean floors in some Walmart locations. It enables customer service bots at Fidelity and interprets X-rays at Zebra Medical. In each case the AI factory treats decision-making as a science. Analytics systematically convert internal and external data into predictions, insights, and choices, which in turn guide and automate operational workflows.

Oddly enough, the AI that can drive the explosive growth of a digital firm often isn't even all that sophisticated. To bring about dramatic change, AI doesn't need to be the stuff of science fiction—indistinguishable from human behavior or simulating human reasoning, a capability sometimes referred to as "strong AI." You need only a computer system to be able to perform tasks traditionally handled by people—what is often referred to as "weak AI."

With weak AI, the AI factory can already take on a range of critical decisions. In some cases it might manage information businesses (such as Google and Facebook). In other cases it will guide how the company builds, delivers, or operates actual physical products (like Amazon's warehouse robots or Waymo, Google's self-driving car service). But in all cases digital decision factories handle some of the most critical processes and operating decisions. Software makes up the core of the firm, while humans are moved to the edge.

Four components are essential to every factory. The first is the data pipeline, the semiautomated process that gathers, cleans, integrates, and safeguards data in a systematic, sustainable, and scalable way. The second is algorithms, which generate predictions about future states or actions of the business. The third is an experimentation platform, on which hypotheses regarding new algorithms are tested to ensure that their suggestions are having the intended effect. The fourth is infrastructure, the systems that embed this process in software and connect it to internal and external users.

Take a search engine like Google or Bing. As soon as someone starts to type a few letters into the search box, algorithms dynamically predict the full search term on the basis of terms that many users have typed in before and

The AI that drives the explosive growth of a firm often isn't even all that sophisticated. AI doesn't need to be the stuff of science fiction, simulating human reasoning.

this particular user's past actions. These predictions are captured in a drop-down menu (the "autosuggest box") that helps the user zero in quickly on a relevant search. Every keystroke and every click are captured as data points, and every data point improves the predictions for future searches. AI also generates the organic search results, which are drawn from a previously assembled index of the web and optimized according to the clicks generated on the results of previous searches. The entry of the term also sets off an automated auction for the ads most relevant to the user's search, the results of which are shaped by additional experimentation and learning loops. Any click on or away from the search query or search results page provides useful data. The more searches, the better the predictions, and the better the predictions, the more the search engine is used.

Removing Limits to Scale, Scope, and Learning

The concept of scale has been central in business since at least the Industrial Revolution. The great Alfred Chandler described how modern industrial firms could reach unprecedented levels of production at much lower unit cost, giving large firms an important edge over smaller rivals. He also highlighted the benefits companies could reap from the ability to achieve greater production scope, or variety. The push for improvement and innovation added a third requirement for firms: learning. Scale, scope, and learning have come to be considered the essential drivers of a firm's operating performance. And for a long time they've been enabled by carefully defined business processes that rely on labor and management to deliver products and services to customers—and that are reinforced by traditional IT systems.

After hundreds of years of incremental improvements to the industrial model, the digital firm is now radically changing the scale, scope, and learning paradigm. AI-driven processes can be scaled up much more rapidly than traditional processes can, allow for much greater scope because they can easily be connected with other digitized businesses, and create incredibly powerful opportunities for learning and improvement—like the ability to produce ever more accurate and sophisticated customer-behavior models and then tailor services accordingly.

In traditional operating models, scale inevitably reaches a point at which it delivers diminishing returns. But we don't necessarily see this with AI-driven models, in which the return on scale can continue to climb to previously unheard-of levels. (See the exhibit "How AI-Driven Companies Can Outstrip Traditional Firms.") Now imagine what happens when an AI-driven firm competes with a traditional firm by serving the same customers with a similar (or better) value proposition and a much more scalable operating model.

We call this kind of confrontation a "collision." As both learning and network effects amplify volume's impact on value creation, firms built on a digital core can overwhelm traditional organizations. Consider the outcome when Amazon collides with traditional retailers, Ant Financial with traditional banks, and Didi and Uber with traditional taxi services. As Clayton Christensen, Michael Raynor, and Rory McDonald argued in "What Is Disruptive Innovation?" (HBR, December 2015), such competitive upsets don't fit the disruption model. Collisions are not caused by a particular innovation in a technology or a business model. They're the result of the emergence of a completely different kind of firm. And they can fundamentally alter industries and reshape the nature of competitive advantage.

Idea in Brief

THE MARKET CHANGE
We're seeing the emergence of a new kind of firm—one in which artificial intelligence is the main source of value creation and delivery.

THE CHALLENGE
The AI-driven operating model is blurring the lines that used to separate industries and is upending the rules of business competition.

THE UPSHOT
For digital start-ups and traditional firms alike, it's essential to understand the revolutionary impact AI has on operations, strategy, and competition.

Note that it can take quite a while for AI-driven operating models to generate economic value anywhere near the value that traditional operating models generate at scale. Network effects produce little value before they reach critical mass, and most newly applied algorithms suffer from a "cold start" before acquiring adequate data. Ant Financial grew rapidly, but its core payment service, Alipay, which had been launched in 2004 by Alibaba, took years to reach its current volume. This explains why executives ensconced in the traditional model have a difficult time at first believing that the digital model will ever catch up. But once the digital operating model really gets going, it can deliver far superior value and quickly overtake traditional firms.

Collisions between AI-driven and traditional firms are happening across industries: software, financial services, retail, telecommunications, media, health care, automobiles, and even agribusiness. It's hard to think of a business that isn't facing the pressing need to digitize its operating model and respond to the new threats.

Rebuilding Traditional Enterprises

For leaders of traditional firms, competing with digital rivals involves more than deploying enterprise software or even building data pipelines, understanding algorithms, and experimenting. It requires rearchitecting the firm's organization and operating model. For a very, very long time, companies have optimized their scale, scope, and learning through greater focus and specialization, which led to the siloed structures that the vast majority of enterprises today have. Generations of information technology didn't change this pattern. For decades, IT was used to enhance the performance of specific functions and organizational units. Traditional enterprise systems often even reinforced silos and the divisions across functions and products.

Silos, however, are the enemy of AI-powered growth. Indeed, businesses like Google Ads and Ant Financial's MyBank deliberately forgo them and are designed to leverage an integrated core of data and a unified, consistent code base. When each silo in a firm has its own data and code, internal development is fragmented, and it's nearly impossible to build connections across the silos or with external business networks or ecosystems. It's also nearly impossible to develop a 360-degree understanding of the customer that both serves and draws from every department and function. So when firms set up a new digital core, they should avoid creating deep organizational divisions within it.

Microsoft's AI Transformation

Microsoft's transformation into an AI-driven firm took years of research but gained steam with the reorganization of its internal IT and data assets, which had been dispersed across the company's various operations. That effort was led by Kurt DelBene, the former head of Microsoft's Office business, who'd left to help fix the U.S. government's HealthCare.gov site before returning to Microsoft in 2015.

There's a reason that CEO Satya Nadella chose someone with product experience to run IT and build the "AI factory" that would be the foundation of the firm's new operating model. "Our product is the process," DelBene told us. "First, we are going to articulate what the vision should be for the systems and processes we support. Second, we're going to be run like a product development team. And we're going to be agile-based." To strengthen that orientation on his team, he brought in handpicked leaders and engineers from the product functions.

Today Core Engineering—as the IT operation is now known—is a showcase for Microsoft's own transformation. Thanks to the group's work, many traditional processes that used to be performed in silos are enabled by one consistent software base residing in Microsoft's Azure cloud. In addition, the team is driving toward a common data architecture across the company. The new, AI-based operating platform connects the sprawling organization with a shared software-component library, algorithm repository, and data catalog, all used to rapidly enable and deploy digital processes across different lines of business.

Beyond increasing productivity and scalability, the AI also helps head off problems. "We leverage AI to know when things are starting to behave in unexpected ways," DelBene says. "The best we could do in the past is react as fast as possible. Now we can preempt things, from bad contracts to cyberbreaches."

You don't have to be a software start-up to digitize critical elements of your business—but you do have to confront silos and fragmented legacy systems, add capabilities, and retool your culture.

While the transition to an AI-driven model is challenging, many traditional firms—some of which we've worked with—have begun to make the shift. In fact, in a recent study we looked at more than 350 traditional enterprises in both service and manufacturing sectors and found that the majority had started building a greater focus on data and analytics into their organizations. Many—including Nordstrom, Vodafone, Comcast, and Visa—had already made important inroads, digitizing and redesigning key components of their operating models and developing sophisticated data platforms and AI capabilities. You don't have to be a software start-up to digitize critical elements of your business—but you do have to confront silos and fragmented legacy systems, add capabilities, and retool your culture. (For a closer look at the key principles that should drive such transformations, see the sidebar "Putting AI at the Firm's Core.")

Fidelity Investments is using AI to enable processes in important areas, including customer service, customer insights, and investment recommendations. Its AI initiatives build on a multiyear effort to integrate data assets into one digital core and redesign the organization around it. The work is by no means finished, but the impact of AI is already evident in many high-value use cases across the company. To take on Amazon, Walmart is rebuilding its operating model around AI and replacing traditional siloed enterprise software systems with an integrated, cloud-based architecture. That will allow Walmart to use its unique data assets in a variety of powerful new applications and automate or enhance a growing number of operating tasks with AI and analytics. At Microsoft, Nadella is betting the company's future on a wholesale transformation of its operating model. (See the sidebar "Microsoft's AI Transformation.")

Rethinking Strategy and Capabilities

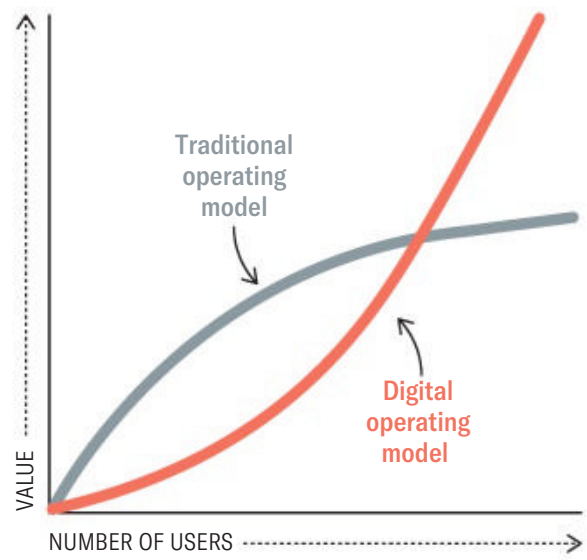
As AI-powered firms collide with traditional businesses, competitive advantage is increasingly defined by the ability to shape and control digital networks. (See "Why Some Platforms Thrive and Others Don't," HBR, January–February 2019.) Organizations that excel at connecting businesses, aggregating the data that flows among them, and extracting its value through analytics and AI will have the upper hand. Traditional network effects and AI-driven learning curves will reinforce each other, multiplying each other's impact. You can see this dynamic in companies such as Google, Facebook, Tencent, and Alibaba, which have become powerful "hub" firms by accumulating data through their many network connections and building the algorithms necessary to heighten competitive advantages across disparate industries.

Meanwhile, conventional approaches to strategy that focus on traditional industry analysis are becoming increasingly ineffective. Take automotive companies. They're facing a variety of new digital threats, from Uber to Waymo, each coming from outside traditional industry boundaries. But if auto executives think of cars beyond their traditional industry context, as a highly connected, AI-enabled service, they can not only defend themselves but also unleash new value—through local commerce opportunities, ads, news and entertainment feeds, location-based services, and so on.

The advice to executives was once to stick with businesses they knew, in industries they understood. But synergies in algorithms and data flows do not respect industry boundaries. And organizations that can't leverage customers and data across those boundaries are likely to be at a big disadvantage. Instead of focusing on industry

How AI-Driven Companies Can Outstrip Traditional Firms

The value that scale delivers eventually tapers off in traditional operating models, but in digital operating models, it can climb much higher.



Putting AI at the Firm's Core

The transition from a traditional firm to an AI-driven organization cannot happen in a skunkworks or be spearheaded by some separate autonomous group. It requires a holistic effort. In our research and our work with a variety of companies, we've come up with five principles that should guide transformations (beyond common best practices for leading change):

One strategy. Rearchitecting a company's operating model means rebuilding each business unit on a new, integrated foundation of data, analytics, and software. This challenging and time-consuming undertaking demands focus and a consistent top-down mandate to coordinate and inspire the many bottom-up efforts involved.

A clear architecture. A new approach based on data, analytics, and AI requires some centralization and a lot of consistency. Data assets should be integrated across a range of applications to maximize their impact. Fragmented data will be virtually impossible to safeguard consistently, especially given privacy and security considerations. If the data isn't all held in centralized repositories, then the organization must at least have an accurate catalog of where the data is, explicit guidelines for what to do with it (and how to protect it), and standards for when and how to store it so that it can be used and reused by multiple parties.

The right capabilities. Though building a base of software, data science, and advanced analytics capabilities will take time, much can be done with a small number of motivated, knowledgeable people. However, many organizations fail to realize that they need to systematically hire a very different kind of talent and set up career paths and incentive systems for those employees.

An agile "product" focus. Building an AI-centric operating model is about taking traditional processes and transforming them into software. Developing a product-focused mentality is essential to getting this done. Like the product managers at any world-class software development project, the IT teams deploying AI-centered applications should have a deep understanding of the use cases they're enabling—a product management orientation that goes well beyond the approach of traditional IT organizations. In the past, IT was largely about keeping old systems working, deploying software updates, protecting against cyberattacks, and running help desks. Developing operating-model software is a different game.

Multidisciplinary governance. The governance of digital assets has become increasingly important and complex and calls for well-thought-out collaboration across disparate disciplines and functions. The challenges of data privacy, algorithmic bias, and cybersecurity are increasing risk and even government intervention and regulation. Governance should integrate a legal and corporate affairs function, which may even be involved in product and technology decisions. AI requires deep thinking about legal and ethical challenges, including careful consideration of what data should be stored and preserved (and what data should not).

analysis and on the management of companies' internal resources, strategy needs to focus on the connections firms create across industries and the flow of data through the networks the firms use.

All this has major implications for organizations and their employees. Machine learning will transform the nature of almost every job, regardless of occupation, income level, or specialization. Undoubtedly, AI-based operating models can exact a real human toll. Several studies suggest that perhaps half of current work activities may be replaced by AI-enabled systems. We shouldn't be too surprised by that. After all, operating models have long been designed to make many tasks predictable and repeatable. Processes for scanning products at checkout, making lattes, and removing hernias, for instance, benefit from standardization and don't require too much human creativity. While AI improvements will enrich many jobs and generate a variety of interesting opportunities, it seems inevitable that they will also cause widespread dislocation in many occupations.

The dislocations will include not only job replacement but also the erosion of traditional capabilities. In almost every setting, AI-powered firms are taking on

While AI improvements will enrich many jobs and generate a variety of interesting opportunities, they will also cause widespread dislocation in many occupations.

highly specialized organizations. In an AI-driven world, the requirements for competition have less to do with specialization and more to do with a universal set of capabilities in data sourcing, processing, analytics, and algorithm development. These new universal capabilities are reshaping strategy, business design, and even leadership. Strategies in very diverse digital and networked businesses now look similar, as do the drivers of operating performance. Industry expertise has become less critical. When Uber looked for a new CEO, the board hired someone who had previously run a digital firm—Expedia—not a limousine services company.

We're moving from an era of core competencies that differ from industry to industry to an age shaped by data and analytics and powered by algorithms—all hosted in the cloud for anyone to use. This is why Alibaba and Amazon are able to compete in industries as disparate as retail and financial services, and health care and credit scoring. These sectors now have many similar technological foundations and employ common methods and tools. Strategies are shifting away from traditional differentiation based on cost, quality, and brand equity and specialized, vertical expertise and toward advantages like business network position, the accumulation of unique data, and the deployment of sophisticated analytics.

The Leadership Challenge

Though it can unleash enormous growth, the removal of operating constraints isn't always a good thing. Frictionless systems are prone to instability and hard to stop once they're in motion. Think of a car without brakes or a skier who can't slow down. A digital signal—a viral meme, for instance—can spread rapidly through networks and can be just about impossible to halt, even for the organization that launched it in the first place or an entity that controls the key hubs in a network. Without friction, a video inciting violence or a phony or manipulative headline can quickly spread to billions of people on a variety of networks, even morphing to optimize click-throughs and downloads. If you have a message to send, AI offers a fantastic way to reach vast numbers of people and personalize that message for them. But the marketer's paradise can be a citizen's nightmare.

Digital operating models can aggregate harm along with value. Even when the intent is positive, the potential downside can be significant. A mistake can expose a large digital network to a destructive cyberattack. Algorithms, if left unchecked, can exacerbate bias and misinformation on a massive scale. Risks can be greatly magnified. Consider the way that digital banks are aggregating consumer savings in an unprecedented fashion. Ant Financial, which now operates one of the largest money market funds in the world, is entrusted with the savings of hundreds of millions of Chinese consumers. The risks that presents are significant, especially for a relatively unproven institution.

Digital scale, scope, and learning create a slew of new challenges—not just privacy and cybersecurity problems, but social turbulence resulting from market concentration, dislocations, and increased inequality. The institutions designed to keep an eye on business—regulatory bodies, for example—are struggling to keep up with all the rapid change.

In an AI-driven world, once an offering's fit with a market is ensured, user numbers, engagement, and revenues can skyrocket. Yet it's increasingly obvious that unconstrained growth is dangerous. The potential for businesses that embrace digital operating models is huge, but the capacity to inflict widespread harm needs to be explicitly considered. Navigating these opportunities and threats will be a real test of leadership for both businesses and public institutions.

HBR Reprint R2001C

Marco Iansiti *is the David Sarnoff Professor of Business Administration at Harvard Business School, where he is the faculty chair for Technology and Transformation. He has advised many companies in the technology sector, including Microsoft, Facebook, and Amazon.* ***Karim R. Lakhani*** *is the Dorothy and Michael Hintze Professor of Business Administration at Harvard Business School and the founder and codirector of the Laboratory for Innovation Science at Harvard. They are the coauthors of the book* Competing in the Age of AI *(Harvard Business Review Press, 2020).*

ORIGINALLY PUBLISHED ON HBR.ORG JULY 18, 2017

The Business of Artificial Intelligence

What it can—and cannot—do for your organization

→ by ERIK BRYNJOLFSSON and ANDREW McAFEE

FOR MORE THAN 250 years the fundamental drivers of economic growth have been technological innovations. The most important of these are what economists call general-purpose technologies—a category that includes the steam engine, electricity, and the internal combustion engine. Each one catalyzed waves of complementary innovations and opportunities. The internal combustion engine, for example, gave rise to cars, trucks, airplanes, chain saws, and lawnmowers, along with big-box retailers, shopping centers, cross-docking warehouses, new supply chains, and, when

Illustration by **MICHELLE D'URBANO**

you think about it, suburbs. Companies as diverse as Walmart, UPS, and Uber found ways to leverage the technology to create profitable new business models.

The most important general-purpose technology of our era is artificial intelligence, particularly machine learning (ML)—that is, the machine's ability to keep improving its performance without humans having to explain exactly how to accomplish all the tasks it's given. Within just the past few years machine learning has become far more effective and widely available. We can now build systems that learn how to perform tasks on their own.

Why is this such a big deal? Two reasons. First, we humans know more than we can tell: We can't explain exactly how we're able to do a lot of things—from recognizing a face to making a smart move in the ancient Asian strategy game of Go. Prior to ML, this inability to articulate our own knowledge meant that we couldn't automate many tasks. Now we can.

Second, ML systems are often excellent learners. They can achieve superhuman performance in a wide range of activities, including detecting fraud and diagnosing disease. Excellent digital learners are being deployed across the economy, and their impact will be profound.

In the sphere of business, AI is poised to have a transformational impact, on the scale of earlier general-purpose technologies. Although it is already in use in thousands of companies around the world, most big opportunities have not yet been tapped. The effects of AI will be magnified in the coming decade, as manufacturing, retailing, transportation, finance, health care, law, advertising, insurance, entertainment, education, and virtually every other industry transform their core processes and business models to take advantage of machine learning. The bottleneck now is in management, implementation, and business imagination.

Like so many other new technologies, however, AI has generated lots of unrealistic expectations. We see business plans liberally sprinkled with references to machine learning, neural nets, and other forms of the technology, with little connection to its real capabilities. Simply calling a dating site "AI-powered," for example, doesn't make it any more effective, but it might help with fundraising. This article will cut through the noise to describe the real potential of AI, its practical implications, and the barriers to its adoption.

What Can AI Do Today?

The term *artificial intelligence* was coined in 1955 by John McCarthy, a math professor at Dartmouth who organized the seminal conference on the topic the following year. Ever since, perhaps in part because of its evocative name, the field has given rise to more than its share of fantastic claims and promises. In 1957 the economist Herbert Simon predicted that computers would beat humans at chess within 10 years. (It took 40.) In 1967 the cognitive scientist Marvin Minsky said, "Within a generation the problem of creating 'artificial intelligence' will be substantially solved." Simon and Minsky were both intellectual giants, but they erred badly. Thus it's understandable that dramatic claims about future breakthroughs meet with a certain amount of skepticism.

Let's start by exploring what AI is already doing and how quickly it is improving. The biggest advances have been in two broad areas: perception and cognition. In the former category some of the most practical advances have been made in relation to speech. Voice recognition is still far from perfect, but millions of people are now using it—think Siri, Alexa, and Google Assistant. The text you are now reading was originally dictated to a computer and transcribed with sufficient accuracy to make it faster than typing. A study by the Stanford computer scientist James Landay and colleagues found that speech recognition is now about three times as fast, on average, as typing on a cell phone. The error rate, once 8.5%, has dropped to 4.9%. What's striking is that this substantial improvement has come not over the past 10 years but just since the summer of 2016.

Image recognition, too, has improved dramatically. You may have noticed that Facebook and other apps now recognize many of your friends' faces in posted photos and prompt you to tag them with their names. An app running on your smartphone will recognize virtually any bird in the wild. Image recognition is even replacing ID cards at corporate headquarters. Vision systems, such as those used in self-driving cars, formerly made a mistake when identifying a pedestrian as often as once in 30 frames (the cameras in these systems record about 30 frames a second); now they err less often than once in 30 million frames. The error rate for recogniz-

Although AI is already in use in thousands of companies around the world, most big opportunities have not yet been tapped.

ing images from a large database called ImageNet, with several million photographs of common, obscure, or downright weird images, fell from higher than 30% in 2010 to about 4% in 2016 for the best systems. (See the exhibit "Puppy or Muffin?")

The speed of improvement has accelerated rapidly in recent years as a new approach, based on very large or "deep" neural nets, was adopted. The ML approach for vision systems is still far from flawless—but even people have trouble quickly recognizing puppies' faces or, more embarrassingly, see their cute faces where none exist.

The second type of major improvement has been in cognition and problem-solving. Machines have already beaten the finest (human) players of poker and Go—achievements that experts had predicted would take at least another decade. Google's DeepMind team has used ML systems to improve the cooling efficiency at data centers by more than 15%, even after they were optimized by human experts. Intelligent agents are being used by the cybersecurity company Deep Instinct to detect malware, and by PayPal to prevent money laundering. A system using IBM technology automates the claims process at an insurance company in Singapore, and a system from Lumidatum, a data science platform firm, offers timely advice to improve customer support. Dozens of companies are using ML to decide which trades to execute on Wall Street, and more and more credit decisions are made with its help. Amazon employs ML to optimize inventory and improve product recommendations to customers. Infinite Analytics developed one ML system to predict whether a user would click on a particular ad, improving online ad placement for a global consumer packaged goods company, and another to improve customers' search and discovery process at a Brazilian online retailer. The first system increased advertising ROI threefold, and the second resulted in a $125 million increase in annual revenue.

Machine learning systems are not only replacing older algorithms in many applications but are now superior at many tasks that were once done best by humans. Although the systems are far from perfect, their error rate—about 5%—on the ImageNet database is at or better than human-level performance. Voice recognition, too, even in noisy environments, is now nearly equal to human performance. Reaching this threshold opens up vast new possibilities for transforming the workplace and the economy. Once AI-based systems surpass human performance at a given task, they are much likelier to spread quickly. For instance, Aptonomy and Sanbot, makers respectively of drones and robots, are using improved vision systems to automate much of the work of security guards. The software company Affectiva, among others, is using them to recognize emotions such as joy, surprise, and anger in focus groups. And Enlitic is one of several deep-learning start-ups that use them to scan medical images to help diagnose cancer.

These are impressive achievements, but the applicability of AI-based systems is still quite narrow. For instance, their remarkable performance on the

Idea in Brief

ON THE HORIZON
Soon, almost every industry will transform its key processes and business models to take advantage of machine learning (ML). But not all expectations surrounding AI are realistic.

THE POSSIBILITIES
Machine-based systems learn over time, and so chances are excellent that AI will reach superhuman levels of performance, especially in the areas of perception and cognition. This means widespread application in tasks from voice and image recognition to forecasting sales, deciding hiring and promotion, and driving a car.

THE LIMITS
Limitations of ML include the risk of data-derived hidden biases, the impossibility of guaranteeing all-case success, and the difficulty of understanding what's going on under the hood—and thus of diagnosing and correcting any errors. And while AI provides us with many answers, it's still humans who pose the questions.

Once AI-based systems surpass human performance at a given task, they are much likelier to spread quickly.

ImageNet database, even with its millions of images, doesn't always translate into similar success "in the wild," where lighting conditions, angles, image resolution, and context may be very different. More fundamentally, we can marvel at a system that understands Chinese speech and translates it into English, but we don't expect such a system to know what a particular Chinese character means—let alone where to eat in Beijing. If someone performs a task well, it's natural to assume that the person has some competence in related tasks. But ML systems are trained to do specific tasks, and typically their knowledge does not generalize. The fallacy that a computer's narrow understanding implies broader understanding is perhaps the biggest source of confusion, and exaggerated claims, about AI's progress. We are far from machines that exhibit general intelligence across diverse domains.

Understanding Machine Learning

The most important thing to understand about ML is that it represents a fundamentally different approach to creating software: The machine learns from examples, rather than being explicitly programmed for a particular outcome. This is an important break from previous practice. For most of the past 50 years, advances in information technology and its applications have focused on codifying existing knowledge and procedures and embedding them in machines. Indeed, the term "coding" denotes the painstaking process of transferring knowledge from developers' heads into a form that machines can understand and execute. This approach has a fundamental weakness: Much of the knowledge we all have is tacit, meaning that we can't fully explain it. It's nearly impossible for us to write down instructions that would enable another person to learn how to ride a bike or to recognize a friend's face.

In other words, we all know more than we can tell. This fact is so important that it has a name: Polanyi's Paradox, for the philosopher and polymath Michael Polanyi, who described it in 1964. Polanyi's Paradox not only limits what we can tell one another but has historically placed a fundamental restriction on our ability to endow machines with intelligence. For a long time that limited the activities that machines could productively perform in the economy.

Machine learning is overcoming those limits. In this second wave of the second machine age, machines built by humans are learning from examples and using structured feedback to solve on their own problems such as Polanyi's classic one of recognizing a face.

Different Flavors of Machine Learning

Artificial intelligence and machine learning come in many flavors, but most of the successes in recent years have been in one category: supervised learning systems, in which the machine is given lots of examples of the correct answer to a particular problem. This process almost always involves mapping from a set of inputs, X, to a set of outputs, Y. For instance, the inputs might be pictures of various animals, and the correct outputs might be labels for those animals: dog, cat, horse. The inputs could also be waveforms from a sound recording and the outputs could be words: "yes," "no," "hello," "good-bye." (See the exhibit "Supervised Learning Systems.")

Successful systems often use a training set of data with thousands or even millions of examples, each of which has been labeled with the correct answer. The system can then be let loose to look at new examples. If the training has gone well, the system will predict answers with a high rate of accuracy.

The algorithms that have driven much of this success depend on an approach called *deep learning,* which uses neural networks. Deep learning algorithms have a significant advantage over earlier generations of ML algorithms: They can make better use of much larger data sets. The old systems would improve as the number of examples in the training data grew, but only up to a point, after which additional data didn't lead to better predictions. According to Andrew Ng, one of the giants of the field, deep neural nets don't seem to level off in this way: More data leads to better and better predictions. Some very large systems are trained by using 36 million examples or more. Of course, working with extremely large data sets requires more and more processing power, which is one reason the very big systems are often run on supercomputers or specialized computer architectures.

Any situation in which you have a lot of data on behavior and are trying to

predict an outcome is a potential application for supervised learning systems. Jeff Wilke, who leads Amazon's consumer business, says that supervised learning systems have largely replaced the memory-based filtering algorithms that were used to make personalized recommendations to customers. In other cases, classic algorithms for setting inventory levels and optimizing supply chains have been replaced by more efficient and robust systems based on machine learning. JPMorgan Chase introduced a system for reviewing commercial loan contracts; work that used to take loan officers 360,000 hours can now be done in a few seconds. And supervised learning systems are now being used to diagnose skin cancer. These are just a few examples.

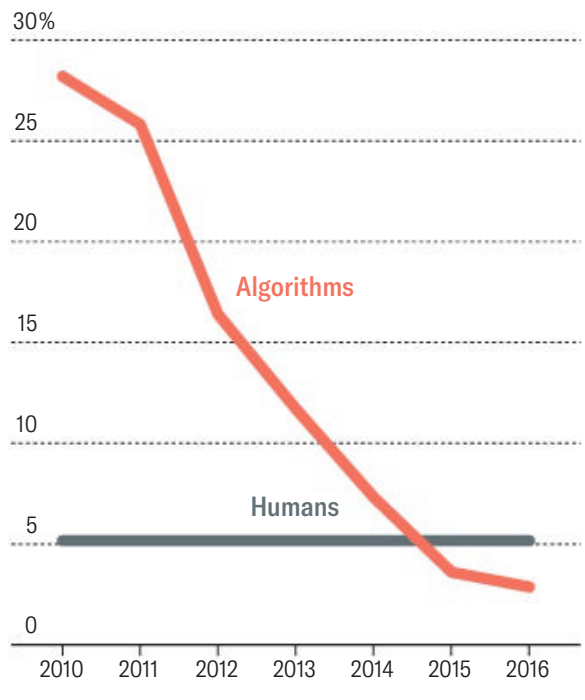
It's comparatively straightforward to label a body of data and use it to train a supervised learner; that's why supervised ML systems are more common than *un*supervised ones, at least for now. Unsupervised learning systems seek to learn on their own. We humans are excellent unsupervised learners: We pick up most of our knowledge of the world (such as how to recognize a tree) with little or no labeled data. But it is exceedingly difficult to develop a successful machine learning system that works this way.

If and when we learn to build robust unsupervised learners, exciting possibilities will open up. These machines could look at complex problems in fresh ways to help us discover patterns—in the spread of diseases, in price moves across securities in a market, in customers' purchase behaviors, and so on—that we are currently unaware of. Such possibilities lead Yann LeCun, the head of AI research at Facebook and a professor at NYU, to compare supervised learning systems to the frosting on the cake and unsupervised learning to the cake itself.

Another small but growing area within the field is *reinforcement learning*. This approach is embedded in systems that have mastered Atari video games and board games like Go. It is also helping to optimize data center power usage and to develop trading strategies for the stock market. Robots created by Kindred use machine learning to identify and sort objects they've never encountered before, speeding up the "pick and place" process in distribution centers for consumer goods. In reinforcement learning systems the programmer specifies the current state of the system and the goal, lists allowable actions, and describes the elements of the environment that constrain the outcomes for each of those actions. Using the allowable actions, the system has to figure out how to get as close to the goal as possible. These systems work well when humans can specify the goal but not necessarily how to get there. For instance, Microsoft used reinforcement learning to select headlines for MSN.com news stories by "rewarding" the system with a higher score when more visitors clicked on the link. The system tried to maximize its score on the basis of the rules its designers gave it. Of course, this means that a reinforcement learning system will optimize for the goal you explicitly reward, not necessarily the goal you really care about (such as lifetime customer value), so specifying the goal correctly and clearly is critical.

Puppy or Muffin? Progress in Image Recognition

Machines have made real strides in distinguishing among similar-looking categories of images.



Source: Electronic Frontier Foundation

KAREN ZACK/@TEENYBISCUIT

Putting Machine Learning to Work

There are three pieces of good news for organizations looking to put ML to use today. First, AI skills are spreading quickly. The world still has not nearly enough data scientists and machine learning experts, but the demand for them is being met by online educational resources as well as by universities. The best of these, including Udacity, Coursera, and fast.ai, do much more than teach introductory concepts; they can actually get smart, motivated students to the point of being able to create industrial-grade ML deployments. In addition to training their own people, interested companies can use online talent platforms such as Upwork, Topcoder, and Kaggle to find ML experts with verifiable expertise.

The second welcome development is that the necessary algorithms and hardware for modern AI can be bought or rented as needed. Google, Amazon, Microsoft, Salesforce, and other companies are making powerful ML infrastructure available via the cloud. The cutthroat competition among these rivals means that companies that want to experiment with or deploy ML will see more and more capabilities available at ever-lower prices over time.

The final piece of good news, and probably the most underappreciated, is that you may not need all that much data to start making productive use of ML. The performance of most machine learning systems improves as they're given more data to work with, so it seems logical to conclude that the company with the most data will win.

Supervised Learning Systems

As two pioneers in the field, Tom Mitchell and Michael I. Jordan, have noted, most of the recent progress in machine learning involves mapping from a set of inputs to a set of outputs. Some examples:

Input X	Output Y	Application
Voice recording	Transcript	Speech recognition
Historical market data	Future market data	Trading bots
Photograph	Caption	Image tagging
Drug chemical properties	Treatment efficacy	Pharma R&D
Store transaction details	Is the transaction fraudulent?	Fraud detection
Recipe ingredients	Customer reviews	Food recommendations
Purchase histories	Future purchase behavior	Customer retention
Car locations and speed	Traffic flow	Traffic lights
Faces	Names	Face recognition

That might be the case if "win" means "dominate the global market for a single application such as ad targeting or speech recognition." But if success is defined instead as significantly improving performance, then sufficient data is often surprisingly easy to obtain.

For example, Udacity cofounder Sebastian Thrun noticed that some of his salespeople were much more effective than others when replying to inbound queries in a chat room. Thrun and his graduate student Zayd Enam realized that their chat room logs were essentially a set of labeled training data—exactly what a supervised learning system needs. Interactions that led to a sale were labeled successes, and all others were labeled failures. Zayd used the data to predict what answers successful salespeople were likely to give in response to certain very common inquiries and then shared those predictions with the other salespeople to nudge them toward better performance. After 1,000 training cycles, the salespeople

Designing and implementing new combinations of technologies, human skills, and capital assets to meet customers' needs requires large-scale creativity and planning. It is a task that machines are not very good at.

had increased their effectiveness by 54% and were able to serve twice as many customers at a time.

The AI start-up WorkFusion takes a similar approach. It works with companies to bring higher levels of automation to back-office processes such as paying international invoices and settling large trades between financial institutions. The reason these processes haven't been automated yet is that they're complicated; relevant information isn't always presented the same way every time ("How do we know what currency they're talking about?"), and some interpretation and judgment are necessary. WorkFusion's software watches in the background as people do their work and uses their actions as training data for the cognitive task of classification ("This invoice is in dollars. This one is in yen. This one is in euros..."). Once the system is confident enough in its classifications, it takes over the process.

Machine learning is driving changes at three levels: tasks and occupations, business processes, and business models. An example of task-and-occupation redesign is the use of machine vision systems to identify potential cancer cells—freeing up radiologists to focus on truly critical cases, to communicate with patients, and to coordinate with other physicians. An example of process redesign is the reinvention of the workflow and layout of Amazon fulfillment centers after the introduction of robots and optimization algorithms based on machine learning. Similarly, business models need to be rethought to take advantage of ML systems that can intelligently recommend music or movies in a personalized way. Instead of selling songs à la carte on the basis of consumer choices, a better model might offer a subscription to a personalized station that predicted and played music a particular customer would like, even if the person had never heard it before.

Note that machine learning systems hardly ever replace the entire job, process, or business model. Most often they complement human activities, which can make their work ever more valuable. The most effective rule for the new division of labor is rarely, if ever, "give all tasks to the machine." Instead, if the successful completion of a process requires 10 steps, one or two of them may become automated while the rest become more valuable for humans to do. For instance, the chat room sales support system at Udacity didn't try to build a bot that could take over all the conversations; rather, it advised human salespeople about how to improve their performance. The humans remained in charge but became vastly more effective and efficient. This approach is usually much more feasible than trying to design machines that can do everything humans can do. It often leads to better, more satisfying work for the people involved and ultimately to a better outcome for customers.

Designing and implementing new combinations of technologies, human skills, and capital assets to meet customers' needs requires large-scale creativity and planning. It is a task that machines are not very good at. That makes being an entrepreneur or a business manager one of society's most rewarding jobs in the age of ML.

Risks and Limits

The second wave of the second machine age brings with it new risks. In particular, machine learning systems often have low "interpretability," meaning that humans have difficulty figuring out how the systems reached their decisions. Deep neural networks may have hundreds of millions of connections, each of which contributes a small amount to the ultimate decision. As a result, these systems' predictions tend to resist simple, clear explanation. Unlike humans, machines are not (yet!) good storytellers. They can't always give a rationale for why a particular applicant was accepted or rejected for a job, or a particular medicine was recommended. Ironically, even as we have begun to overcome Polanyi's Paradox, we're facing a kind of reverse version: Machines know more than they can tell us.

This creates three risks. First, the machines may have hidden biases, derived not from any intent of the designer but from the data provided to train the system. For instance, if a system learns which job applicants to accept for an interview by using a data set of decisions made by human recruiters in the past, it may inadvertently learn to perpetuate their racial, gender, ethnic, or other biases. Moreover, these biases may not appear as an explicit rule but, rather, be embedded in subtle interactions among the thousands of factors considered.

A second risk is that, unlike traditional systems built on explicit logic rules, neural network systems deal with statistical truths rather than literal truths. That can make it difficult, if not

impossible, to prove with complete certainty that the system will work in all cases—especially in situations that weren't represented in the training data. Lack of verifiability can be a concern in mission-critical applications, such as controlling a nuclear power plant, or when life-or-death decisions are involved.

Third, when the ML system does make errors, as it almost inevitably will, diagnosing and correcting exactly what's going wrong can be difficult. The underlying structure that led to the solution can be unimaginably complex, and the solution may be far from optimal if the conditions under which the system was trained change.

While all these risks are very real, the appropriate benchmark is not perfection but the best available alternative. After all, we humans, too, have biases, make mistakes, and have trouble explaining truthfully how we arrived at a particular decision. The advantage of machine-based systems is that they can be improved over time and will give consistent answers when presented with the same data.

Does that mean there is no limit to what artificial intelligence and machine learning can do? Perception and cognition cover a great deal of territory—from driving a car to forecasting sales to deciding whom to hire or promote. We believe the chances are excellent that AI will soon reach superhuman levels of performance in most or all of these areas. So what *won't* AI and ML be able to do?

We sometimes hear "Artificial intelligence will never be good at assessing emotional, crafty, sly, inconsistent human beings—it's too rigid and impersonal for that." We don't agree. ML systems like those at Affectiva are already at or beyond human-level performance in discerning a person's emotional state on the basis of tone of voice or facial expression. Other systems can infer when even the world's best poker players are bluffing well enough to beat them at the amazingly complex game Heads-Up No-Limit Texas Hold'em. Reading people accurately is subtle work, but it's not magic. It requires perception and cognition—exactly the areas in which ML is currently strong and getting stronger all the time.

A great place to start a discussion of the limits of AI is with Pablo Picasso's observation about computers: "But they are useless. They can only give you answers." They're actually far from useless, as ML's recent triumphs show, but Picasso's observation still provides insight. Computers are devices for answering questions, not for posing them. That means entrepreneurs, innovators, scientists, creators, and other kinds of people who figure out what problem or opportunity to tackle next, or what new territory to explore, will continue to be essential.

Similarly, there's a huge difference between passively assessing someone's mental state or morale and actively working to change it. ML systems are getting quite good at the former but remain well behind us at the latter. We humans are a deeply social species; other humans, not machines, are best at tapping into social drives such as compassion, pride, solidarity, and shame in order to persuade, motivate, and inspire. In 2014 the TED Conference and the XPrize Foundation announced an award for "the first artificial intelligence to come to this stage and give a TED Talk compelling enough to win a standing ovation from the audience." We doubt the award will be claimed anytime soon.

We think the biggest and most important opportunities for human smarts in this new age of superpowerful ML lie at the intersection of two areas: figuring out what problems to work on next, and persuading a lot of people to tackle them and go along with the solutions. This is a decent definition of leadership, which is becoming much more important in the second machine age.

The status quo of dividing up work between minds and machines is falling apart very quickly. Companies that stick with it are going to find themselves at an ever-greater competitive disadvantage compared with rivals who are willing and able to put ML to use in all the places where it is appropriate and who can figure out how to effectively integrate its capabilities with humanity's.

A time of tectonic change in the business world has begun, brought on by technological progress. As was the case with steam power and electricity, it's not access to the new technologies themselves, or even to the best technologists, that separates winners from losers. Instead, it's innovators who are open-minded enough to see past the status quo and envision very different approaches, and savvy enough to put them into place. One of machine learning's greatest legacies may well be the creation of a new generation of business leaders.

While all the risks of AI are very real, the appropriate benchmark is not perfection but the best available alternative.

In our view, artificial intelligence, especially machine learning, is the most important general-purpose technology of our era. The impact of these innovations on business and the economy will be reflected not only in their direct contributions but also in their ability to enable and inspire complementary innovations. New products and processes are being made possible by better vision systems, speech recognition, intelligent problem-solving, and many other capabilities that machine learning delivers.

Some experts have gone even further. Gil Pratt, who now heads the Toyota Research Institute, has compared the current wave of AI technology to the Cambrian explosion 500 million years ago that birthed a tremendous variety of new life forms. Then as now, one of the key new capabilities was vision. When animals first gained this capability, it allowed them to explore the environment far more effectively; that catalyzed an enormous increase in the number of species, both predators and prey, and in the range of ecological niches that were filled. Today as well we expect to see a variety of new products, services, processes, and organizational forms and also numerous extinctions. There will certainly be some weird failures along with unexpected successes.

Although it is hard to predict exactly which companies will dominate in the new environment, a general principle is clear: The most nimble and adaptable companies and executives will thrive. Organizations that can rapidly sense and respond to opportunities will seize the advantage in the AI-enabled landscape. So the successful strategy is to be willing to experiment and learn quickly. If managers aren't ramping up experiments in the area of machine learning, they aren't doing their job. Over the next decade, AI won't replace managers, but managers who use AI will replace those who don't.

HBR Reprint H03QXY
Big Idea Package Reprint BG1704

Erik Brynjolfsson *is a professor and senior fellow at the Stanford Institute for Human-Centered AI (HAI), the director of the Stanford Digital Economy Lab, and a research associate at NBER. He is the author or coauthor of nine books, including* The Second Machine Age *and* Machine, Platform, Crowd. ***Andrew McAfee*** *is the cofounder of the Initiative on the Digital Economy in the MIT Sloan School of Management. He is the author of* Enterprise 2.0 *and a coauthor of* The Second Machine Age.

Quick Takes

1. How AI Will Change Strategy: A Thought Experiment

→ by AJAY AGRAWAL, JOSHUA GANS, and AVI GOLDFARB

HOW WILL AI change strategy? That's the single most common question corporate executives ask the three of us, and it's not trivial to answer. AI is fundamentally a prediction technology. As advances in AI make prediction cheaper, economic theory dictates that we'll use prediction more frequently and widely, and the value of complements to prediction—like human judgment—will rise. But what does all this mean for strategy?

Here's a thought experiment we've been using to answer that question. Most people are familiar with shopping at Amazon. As with most online retailers, you visit Amazon's website, shop for items, place them in your "basket," pay for them, and then Amazon ships them to you. Right now, Amazon's business model is shopping-then-shipping.

Most shoppers have noticed Amazon's recommendation engine while they shop—it suggests items that its AI predicts you will want to buy. At present, Amazon's AI does a reasonable job, considering the millions of items offered. However, it is far from perfect. In our case, the AI accurately predicts what we want to buy about 5% of the time. In other words, we actually purchase about one out of every 20 items it recommends. Not bad!

Now for the thought experiment. Imagine that the Amazon AI collects more information about us: In addition to our searching and purchasing behavior on the website, it also collects other data it finds online, including social media, as well as offline, such as our shopping behavior at Whole Foods. It knows not only what we buy but also what time we go to the store, which location we shop at, how we pay, and more.

Now, imagine that the AI uses that data to improve its predictions. We think of this sort of improvement as akin to turning up the volume knob on a speaker dial. But rather than volume, you're turning up the AI's prediction accuracy. What happens to Amazon's strategy as the company's data scientists, engineers, and machine-learning experts work tirelessly to dial up the accuracy on the prediction machine?

At some point, as they turn the knob, the AI's prediction accuracy crosses a threshold such that it becomes in Amazon's interest to change its business model. The prediction becomes sufficiently accurate that it becomes more profitable for Amazon to ship you the goods that it predicts you will want rather than wait for you to order them. Every week, Amazon ships you boxes of items it predicts you will want, and then you shop in the comfort

SEAN GLADWELL/GETTY IMAGES

Turning the dial on the prediction machine has a significant impact on strategy.

and convenience of your own home by choosing the items you wish to keep.

This approach offers two benefits to Amazon. First, the convenience of predictive shipping makes it much less likely that you will purchase the items from a competing retailer as the products are conveniently delivered to your home before you buy them elsewhere. Second, predictive shipping nudges you to buy items that you were considering purchasing but might not have gotten around to. In both cases, Amazon gains a higher share-of-wallet. Turning the prediction dial up far enough changes Amazon's business model from shopping-then-shipping to shipping-then-shopping.

Of course, shoppers would not want to deal with the hassle of returning all the items they don't want. So, Amazon would invest in infrastructure for the product returns—perhaps a fleet of delivery-style trucks that do pickups once a week, conveniently collecting items that customers don't want.

If this is a better business model, then why hasn't Amazon done it already? Well, the company may be working on it. But if it were implemented today, the cost of collecting and handling returned items would outweigh the increase in revenue from a greater share-of-wallet. For example, today we would return 95% of the items it ships to us. That is annoying for us and costly for Amazon. The prediction isn't good enough for Amazon to adopt the new model.

That said, one can imagine a scenario where Amazon adopts the new strategy even *before* the prediction accuracy is good enough to make it profitable because the company *anticipates* that at some point it will be profitable. By launching sooner, Amazon's AI will get more data sooner and improve faster. Amazon realizes that the sooner it gets started, the harder it will be for competitors to catch up. Better predictions will attract more shoppers, more shoppers will generate more data to train the AI, more data will lead to better predictions, and so on, creating a virtuous circle. In other words, there are increasing returns to AI, and thus the timing of adopting this kind of strategy matters. Adopting too early could be costly, but adopting too late could be fatal.

The key insight here is that turning the dial on the prediction machine has a significant impact on strategy. In this example, it shifts Amazon's business model from shopping-then-shipping to shipping-then-shopping, generates the incentive to vertically integrate into operating a product-returns service (including a fleet of trucks), and accelerates the timing of investment due to first-mover advantage from increasing returns. All this is due to the single act of turning the dial on the prediction machine.

Most readers are familiar with the outcome of companies like Blockbuster and Borders that underestimated how quickly the online consumer behavior dial would turn in the context of online shopping and the digital distribution of goods and services. Perhaps they were lulled into complacency by the initially slow adoption rate of this technology in the early days of the commercial internet (1995–1998).

Today, in the case of AI, some companies are making early bets anticipating that the dial on the prediction machine will start turning faster once it gains momentum. Most people are aware of Google's 2014 acquisition of DeepMind—over $500 million for a company that had generated negligible revenue but had developed an AI that could play certain Atari games at a superhuman performance level. Perhaps fewer readers know that more-traditional companies are also making bets on the pace the dial will turn. In 2016 GM paid over $1 billion to acquire AI start-up Cruise Automation, and in 2017 Ford invested $1 billion in AI start-up Argo AI, and John Deere paid over $300 million to acquire AI start-up Blue River Technology—all three start-ups had generated negligible revenue relative to the price at the time of purchase. GM, Ford, and John Deere are each betting on an exponential speedup of AI performance and, at those prices, anticipating a significant impact on their business strategies.

Strategists face two issues in light of all this. First, they must invest in developing a better understanding of how fast and how far the dial on their prediction machines will turn for their sector and applications. Second, they must invest in developing a thesis about the strategy options created by the shifting economics of their business that result from turning the dial, similar to the thought experiment we considered for Amazon.

So, the overarching theme for initiating an AI strategy? Close your eyes, imagine putting your fingers on the dial of your prediction machine,

and, in the immortal words of Spinal Tap, turn it to 11.

HBR Reprint H03XDI

Originally published on HBR.org October 3, 2017

The ideas here are adapted from our book Prediction Machines: The Simple Economics of Artificial Intelligence *(Harvard Business Review Press, 2018).*

Ajay Agrawal *is the Geoffrey Taber Chair in Entrepreneurship and Innovation at the University of Toronto's Rotman School of Management. He is the founder of the Creative Destruction Lab, cofounder of The Next AI, and cofounder of Sanctuary. He is a coauthor of* Prediction Machines: The Simple Economics of Artificial Intelligence *(Harvard Business Review Press, 2018).* ***Joshua Gans*** *is the Jeffrey S. Skoll Chair in Technical Innovation and Entrepreneurship at the Rotman School of Management, University of Toronto, and chief economist at the Creative Destruction Lab. He is the author of* The Disruption Dilemma *(MIT Press, 2016) and a coauthor of* Prediction Machines: The Simple Economics of Artificial Intelligence. ***Avi Goldfarb*** *is the Rotman Chair in Artificial Intelligence and Healthcare at the Rotman School of Management, University of Toronto. He is also the chief data scientist at the Creative Destruction Lab and a coauthor of* Prediction Machines: The Simple Economics of Artificial Intelligence.

PAUL BRADBURY/GETTY IMAGES

2. Why Companies That Wait to Adopt AI May Never Catch Up

→ by VIKRAM MAHIDHAR and THOMAS H. DAVENPORT

WHILE SOME COMPANIES—most large banks, Ford and GM, Pfizer, and virtually all tech firms—are aggressively adopting AI, many are not. Instead they are waiting for the technology to mature and for expertise in AI to become more widely available. They are planning to be "fast followers"—a strategy that has worked with most information technologies.

We think this is a bad idea. It's true that some technologies need further development, but some (like traditional machine learning) are quite mature and have been available in some form for decades. Even more recent technologies, like deep learning, are based on research that took place

Interaction learning is a critical step for organizations to understand how the AI system interacts with its ecosystem.

in the 1980s. New research is being conducted all the time, but the mathematical and statistical foundations of current AI are well established.

System Development Time

Beyond the technical maturity issue, there are several other problems with the idea that companies will be able to adopt quickly once technologies are more capable. First, there is the time required to develop AI systems. Such systems will probably add little value to your business if they are completely generic, so time is required to tailor and configure them to your business and the specific knowledge domain within it. If the AI you are adopting employs machine learning, you will have to round up a substantial amount of training data. If it manipulates language—as in natural language processing applications—it can be even more difficult to get systems up and running. There is a lot of taxonomy and local knowledge that needs to be incorporated into the AI system—similar to the old "knowledge engineering" activity for expert systems. AI of this type is not just a software coding problem; it is a knowledge coding problem. It takes time to discover, disambiguate, and deploy knowledge.

Particularly if your knowledge domain has not already been modeled by your vendor or consultant, it will typically require many months to architect. This is especially true for complex knowledge domains. For example, Memorial Sloan Kettering Cancer Center has been working with IBM to use Watson to treat certain forms of cancer for over six years, and the system still isn't ready for broad use despite availability of high-quality talent in cancer care and AI. There are several domains and business problems for which the requisite knowledge engineering is available. However, it still needs to be manipulated to a company's specific business context.

Integration Time

Even after your AI systems have been built, they need to be integrated into your organization. Unless you are employing some AI capabilities that are embedded within existing packaged application systems that your company already uses (for example, Salesforce Einstein features within your CRM system), the fit with your business processes and IT architecture will require significant planning and time for adaptation. The transition from pilots and prototypes to production systems for AI can be difficult and time-consuming.

Even if your organization is skilled at moving pilots and prototypes into production, you will also have to reengineer the business processes to have full impact on your business and industry. In most cases AI supports individual tasks, not entire business processes, so you will have to redesign business processes and new human tasks around it. If you want to affect customer engagement, for example, you will need to develop or adapt multiple AI applications and tasks that relate to different aspects of marketing, sales, and service relationships.

Human Interactions with AI Time

Finally, there are the human challenges of AI to overcome. Very few AI systems are fully autonomous but instead are focused on augmentation of and by human workers. New AI systems typically mean new roles and skills for the humans who work alongside them, and it will typically require considerable time to retrain workers on the new process and system. For example, investment advice companies providing "robo-advice" to their customers have often attempted to get human advisers to shift their focus to "behavioral finance" or to offer advice and "nudges" to encourage wise decisions and actions in investing. But this sort of skill is quite different from providing advice about what stocks and bonds to buy and will take some time to inculcate.

Even if the goal for an AI system is to be fully autonomous, some period of time in augmentation mode will most likely be necessary. During this period, a critical piece of machine learning occurs through interaction between the system and its human users and observers. Called interaction learning, this is a critical step for organizations to understand how the system interacts with its ecosystem. They can often gather new data sets and begin to bake them into algorithms during this period—which often takes months or years.

Governance Time for AI Applications

While AI systems are geared to provide exponential scale and predictions, they

It may take a long time to develop and fully implement AI systems, and there are few, if any, shortcuts.

will need a much broader governing approach than the classic controls- and test-driven approach. The efficacy of AI algorithms decays over time because they are built on the basis of historical data and recent business knowledge. The algorithms can be updated as the machine learns from patterns in new data, but they will need to be monitored by subject-matter experts to ensure the machine is interpreting the change in business context correctly. Algorithms will also have to be continuously monitored for bias. For instance, if an AI system is trained to create product recommendations on the basis of customer demographics and the demographics change dramatically in new data, it may provide biased recommendations.

Governance will also include watching for customer fraud. As the systems become smart, so will the users. They may try to game the systems with fraudulent data and activities. Monitoring and preventing this will require sophisticated instrumentation and human monitoring in the context of your business.

Winners Take All

It may, then, take a long time to develop and fully implement AI systems, and there are few, if any, shortcuts to the necessary steps. Once they have been successfully undertaken, scaling—particularly if the company has a plentiful supply of data and the knowledge engineering is mastered—can be very rapid. By the time a late adopter has done all the necessary preparation, earlier adopters will have taken considerable market share—they'll be able to operate at substantially lower costs with better performance. In short, the winners may take all and late adopters may never catch up. Think, for example, of the learning and capability that a company like Pfizer—which has, according to one of the leaders of the company's Analytics and AI Lab, more than 150 AI projects underway—has already accumulated. Tech companies like Alphabet have even more learning; that company had 2,700 AI projects underway as far back as 2015.

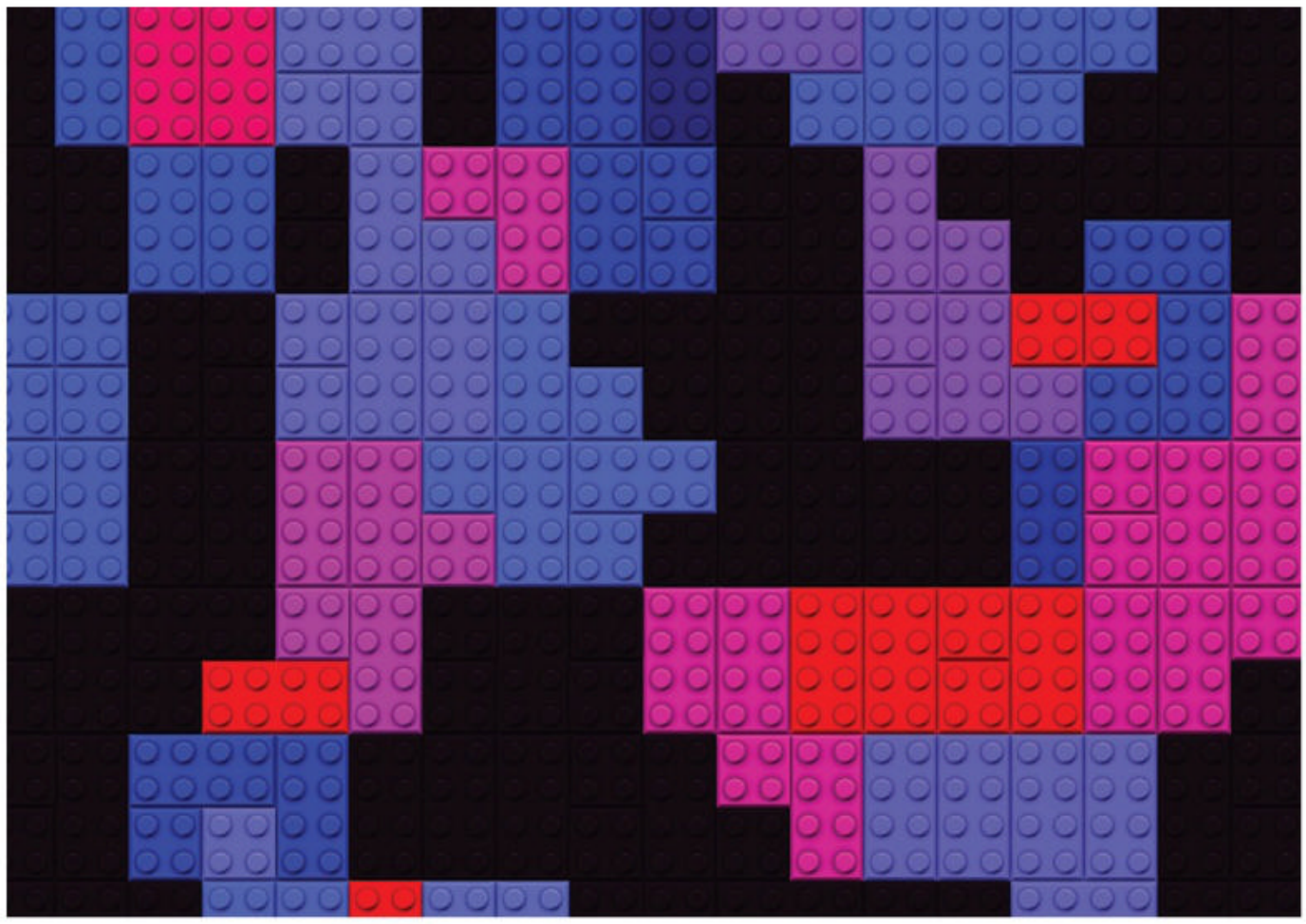
Admittedly, some steps can be accelerated by waiting if a company is willing to compromise its unique knowledge and ways of conducting business. Vendors are developing a vast variety of knowledge graphs and models that use techniques ranging from natural language processing to computer vision. If one exists for your industry or business problem, and you're willing to adopt it with little modification, that will speed up the process of AI adoption. But you may lose your distinctive competence or competitive advantage if you do not tweak it to fit your context and build everything around it.

The obvious implication is that if you want to be successful with AI and think there may be a threat from AI-driven competitors or new entrants, you should start learning now about how to adapt it to your business across multiple different applications and AI methods. Some leading companies have created a centralized AI group to do this at scale. Such central groups focus on framing the problems, proving out the business hypothesis, modularizing the AI assets for reusability, creating techniques to manage the data pipeline, and training across businesses. One other possibility may be to acquire a start-up that has accumulated substantial AI capabilities, but there will still be the need to adapt those capabilities to your business. In short, you should get started now if you haven't already, and hope that it's not too late.

HBR Reprint H04OK4

Originally published on HBR.org December 6, 2018

Vikram Mahidhar *is an operating partner and the head of data and digital transformation at Apollo. He was previously a senior vice president at Genpact. He was also an entrepreneur and built the AI Solutions business at Rage Framework (acquired by Genpact).* ***Thomas H. Davenport*** *is the President's Distinguished Professor of Information Technology and Management at Babson College, a research fellow at the MIT Initiative on the Digital Economy, and a senior adviser at Deloitte Analytics. He is the author of over a dozen management books, most recently* Only Humans Need Apply: Winners and Losers in the Age of Smart Machines *(Harper Business, 2016), with coauthor Julia Kirby, and* The AI Advantage: How to Put the Artificial Intelligence Revolution to Work *(MIT Press, 2018).*



OXYGEN/GETTY IMAGES

3. Why AI That Teaches Itself to Achieve a Goal Is the Next Big Thing

→ by KATHRYN HUME and MATTHEW E. TAYLOR

LEE SEDOL, a world-class Go champion, was flummoxed by the 37th move that DeepMind's AlphaGo made in the second match of the famous 2016 series—so flummoxed that it took him nearly 15 minutes to formulate a response. The move was strange to other experienced Go players as well, with one commentator suggesting it was a mistake. In fact, it was a canonical example of an AI algorithm learning something that seemed to go beyond just pattern recognition in data—learning something strategic and even creative. Indeed, beyond just feeding the algorithm past examples of Go champions playing games, DeepMind developers trained AlphaGo by having it play many millions of matches against itself. During these matches, the system had the chance to explore new moves and strategies, and then evaluate if they improved performance. Through all this trial and error, it discovered a way to play the game that surprised even the best players in the world.

If this kind of AI with creative capabilities seems different from the chatbots and predictive models most businesses end up with when they apply machine learning, that's because it is. Instead of machine learning that uses historical data to generate predictions, game-playing systems like AlphaGo use reinforcement learning—a mature machine-learning technology that's good at optimizing tasks. To do so, an agent takes a series of actions over time, and each action is informed by the outcome of the previous ones. Put simply, it works by trying different approaches and latching on to—reinforcing—the ones that seem to work better than the others. With enough trials, you can reinforce your way to beating your current best approach and discover a new best way to accomplish your task.

Despite its demonstrated usefulness, however, reinforcement learning is mostly used in academia and niche areas like video games and robotics. Companies such as Netflix, Spotify, and Google have started using it, but most businesses lag behind. Yet opportunities are everywhere. In fact, any time you have to make

decisions in sequence—what AI practitioners call sequential decision tasks—there is a chance to deploy reinforcement learning.

Consider the many real-world problems that require deciding how to act over time, where there is something to maximize (or minimize), and where you're never explicitly given the correct solution. For example:

- How should you route data traffic to different servers or decide what servers to power down in a data center?
- When building a molecule in simulation to develop a breakthrough drug, how do you determine which reagent to add next?
- If you want to sell a large amount of stock, how do you carefully sell small orders throughout a day to minimize the amount that the stock price drops?

If you're a company leader, there are probably many processes you'd like to automate or optimize but that are too dynamic or have too many exceptions and edge cases to program into software. Through trial and error, reinforcement learning algorithms can learn to solve even the most dynamic optimization problems—opening up new avenues for automation and personalization in quickly changing environments.

How Top Companies Are Using This Breed of AI to Solve Tough Problems

Company	Application	Sector	Inputs	Actions	Objective
Royal Bank of Canada	Trade execution platform for multiple strategies	Financial services	200-plus market-related data inputs	Sell, buy, hold stocks	To trade as close as possible to VWAP, a common price metric
Netflix	Test schedules for business partner devices	Technology	Historical test and device performance information	Which test to do next	Minimize device failure
Spotify	Recommendation engine	Entertainment	Previous songs liked/disliked/not played	Which songs to put in your playlist	Maximize user listening time
JPMorgan Chase	Financial derivatives risk and pricing calculations	Financial services	Historical market data	Price and sell a financial product	Maximize future cash flows of an investment portfolio
Google	Data center cooling	Technology	Temperature/air pressure	Turn on fan; add water to air unit	Control temperature and reduce energy usage
DiDi	Order dispatching	Ride hailing	Number of idle vehicles, number of orders, location, destination	Match driver to passenger	Minimize pickup time and maximize revenue

Note: Details for use cases can be found in published papers, but we could not verify if they are used in production applications.

What Reinforcement Learning Can Do

Many businesses think of machine-learning systems as "prediction machines" and apply algorithms to forecast things like cash flow or customer attrition on the basis of data such as transaction patterns or website analytics behavior. These systems tend to use what's called supervised machine learning. With supervised learning, you typically make a prediction: For example, the stock will likely go up by four points in the next six

Reinforcement learning algorithms can learn to solve even the most dynamic optimization problems.

hours. Then, after you make that prediction, you're given the actual answer: The stock went up by three points. The system learns by updating its mapping between input data—like past prices of the same stock and perhaps of other equities and indicators—and output prediction to better match the actual answer, which is called the ground truth.

With reinforcement learning, however, there's no correct answer to learn from. Reinforcement learning systems produce actions, not predictions—they'll suggest the action most likely to maximize (or minimize) a metric. You can observe only how well you did on a particular task and whether it was done faster or more efficiently than before. Because these systems learn through trial and error, they work best when they can rapidly try an action (or sequence of actions) and get feedback—a stock market algorithm that takes hundreds of actions per day is a good use case; optimizing customer lifetime value over the course of five years, with only irregular interaction points, is not. Significantly, because of how they learn, they don't need mountains of historical data—they'll experiment and create their own data along the way.

They can therefore be used to automate a process, like placing items into a shipping container with a robotic arm, or to optimize a process, like deciding when and through what channel to contact a client who missed a payment, with the highest recouped revenue and lowest expended effort. In either case, designing the inputs, actions, and rewards the system uses is the key—it will optimize exactly what you encode it to optimize and doesn't do well with any ambiguity.

Google's use of reinforcement learning to help cool its data centers is a good example of how this technology can be applied. Servers in data centers generate a lot of heat, especially when they're in close proximity to one another, and overheating can lead to IT performance issues or equipment damage. In this use case, the input data is various measurements about the environment, like air pressure and temperature. The actions are fan speed (which controls air flow) and valve opening (the amount of water used) in air-handling units. The system includes some rules to follow safe operating guidelines, and it sequences how air flows through the center to keep the temperature at a specified level while minimizing energy usage. The physical dynamics of a data center environment are complex and constantly changing; a shift in the weather impacts temperature and humidity, and each physical location often has a unique architecture and setup. Reinforcement learning algorithms can pick up on nuances that would be too hard to describe with formulas and rules.

Reinforcement Learning at Work

Here at Borealis AI, we partnered with Royal Bank of Canada's Capital Markets business to develop a reinforcement learning–based trade execution system called Aiden. Aiden's objective is to execute a customer's stock order (to buy or sell a certain number of shares) within a specified time window, seeking prices that minimize loss relative to a specified benchmark. This becomes a sequential decision task because of the detrimental market impact of buying or selling too many shares at once: The task is to sequence actions throughout the day to minimize price impact.

The stock market is dynamic, and the performance of traditional algorithms (the rules-based algorithms traders have used for years) can vary when today's market conditions differ from yesterday's. We felt this was a good reinforcement learning opportunity—it had the right balance between clarity and dynamic complexity. We could clearly enumerate the different actions Aiden could take and the reward we wanted to optimize (minimize the difference between the prices Aiden achieved and the market volume–weighted average price benchmark). The stock market moves fast and generates a lot of data, giving the algorithm quick iterations to learn.

We let the algorithm do just that through countless simulations before launching the system live to the market. Ultimately, Aiden proved able to perform well during some of the more volatile market periods during the beginning of the Covid-19 pandemic—conditions that are particularly tough for predictive AIs. It was able to adapt to the changing environment while continuing to stay close to its benchmark target.

How to Spot an Opportunity for Reinforcement Learning

How can you tell if you're overlooking a problem that

Reinforcement learning systems produce actions, not predictions.

reinforcement learning might be able to fix? Here's where to start:

Make a list. Create an inventory of business processes that involve a sequence of steps and clearly state what you want to maximize or minimize. Focus on processes with dense, frequent actions and opportunities for feedback, and avoid processes with infrequent actions and where it's difficult to observe which worked best to collect feedback. Getting the objective right will most likely require iteration.

Consider other options. Don't start with reinforcement learning if you can tackle a problem with other machine-learning or optimization techniques. Reinforcement learning is helpful when you lack sufficient historical data to train an algorithm. You need to explore options (and create data along the way).

Be careful what you wish for. If you do want to move ahead, domain experts should closely collaborate with technical teams to help design the inputs, actions, and rewards. For inputs, seek the smallest set of information you could use to make a good decision. For actions, ask how much flexibility you want to give the system; start simple and later expand the range of actions. For rewards, think carefully about the outcomes—and be careful to avoid falling into the traps of considering one variable in isolation or opting for short-term gains with long-term pains.

Ask whether it's worth it. Will the possible gains justify the costs for development? Many companies need to make digital transformation investments to have the systems and dense data-generating business processes in place to really make reinforcement learning systems useful. To answer whether the investment will pay off, technical teams should take stock of computational resources to ensure you have the compute power required to support trials and allow the system to explore and identify the optimal sequence. (They may want to create a simulation environment to test the algorithm before releasing it live.) On the software front, if you're planning to use a learning system for customer engagement, you need to have a system that can support A/B testing. This is critical to the learning process, as the algorithm needs to explore different options before it can latch on to which one works best. Finally, if your technology stack can only release features universally, you probably need to upgrade before you start optimizing.

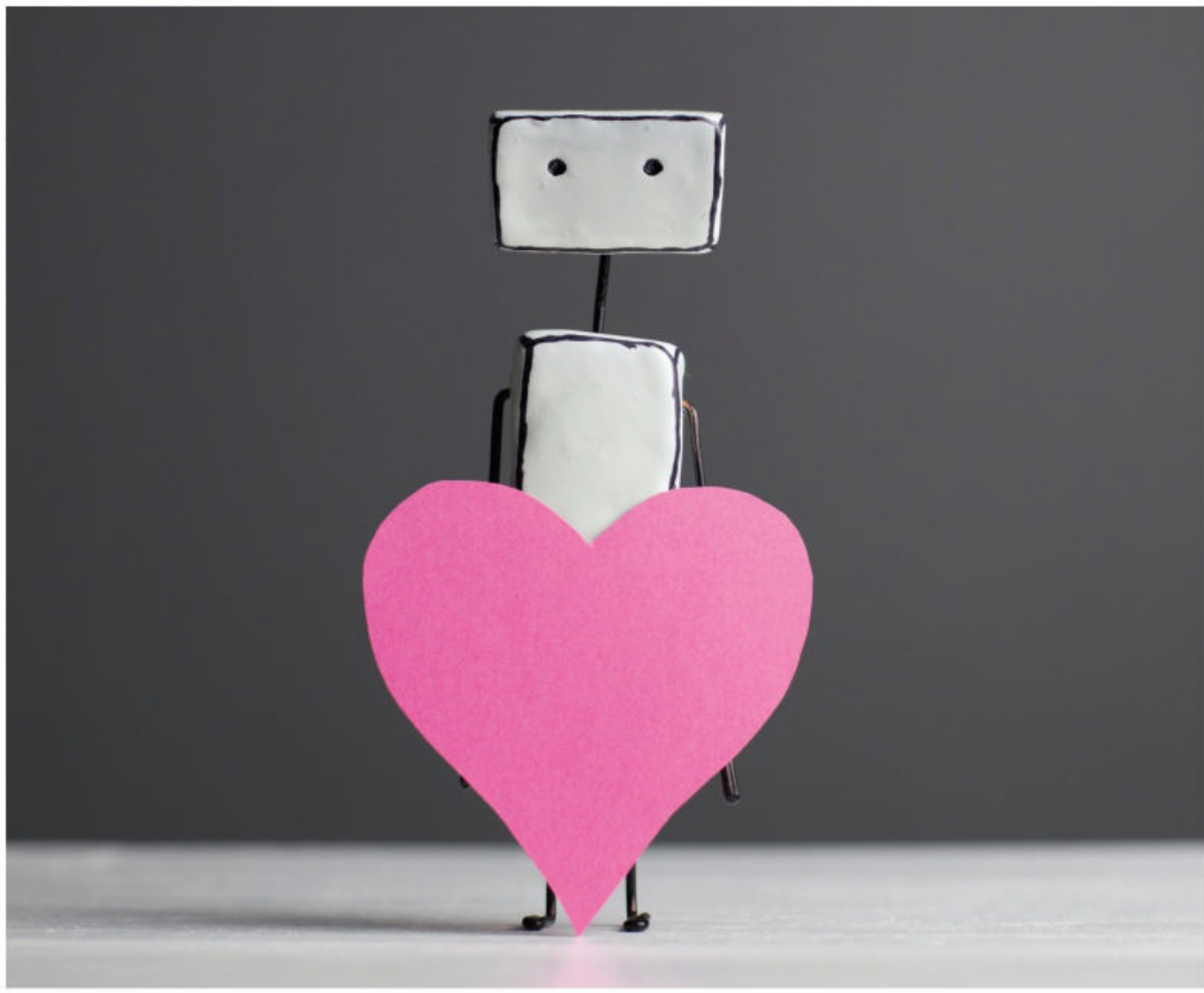
Prepare to be patient. And last but not least, as with many learning algorithms, you have to be open to errors early on while the system learns. It won't find the optimal path from day one, but it will get there in time—and potentially find surprising, creative solutions beyond human imagination when it does.

While reinforcement learning is a mature technology, it's only now starting to be applied in business settings. The technology shines when used to automate or optimize business processes that generate dense data and where there could be unanticipated changes you couldn't capture with formulas or rules. If you can spot an opportunity, and either lean on an in-house technical team or partner with experts in the space, there's a window to apply this technology to outpace your competition.

HBR Reprint H06BMZ

Originally published on HBR.org April 21, 2021

Kathryn Hume *is the interim head of Borealis AI, the machine-learning research lab for the Royal Bank of Canada. Prior to joining Borealis, she held leadership positions at Integrate.ai and Fast Forward Labs, where she helped over 50* Fortune *500 organizations develop and implement AI programs. She has taught courses on digital transformation and legal ethics at the business and law schools at Harvard, MIT, the University of Toronto, and the University of Calgary.* ***Matthew E. Taylor*** *is an associate professor of computing science at the University of Alberta, where he directs the Intelligent Robot Learning Lab, and is a fellow and fellow-in-residence at the Amii (Alberta Machine Intelligence Institute). His current research interests include fundamental improvements to reinforcement learning, the application of reinforcement learning to real-world problems, and human-AI interaction. His upcoming book* Reinforcement Learning Applications for Real-World Data, *with coauthors Phil Osborne and Kajal Singh, is aimed at practitioners without degrees in machine learning and has an expected release date of fall 2021.*



BY WUNDERFOOL/GETTY IMAGES

4. Three Ways AI Is Getting More Emotional

→ by SOPHIE KLEBER

IN JANUARY 2018 Annette Zimmermann, vice president of research at Gartner, proclaimed: "By 2022, your personal device will know more about your emotional state than your own family." Just two months later, a landmark study from the University of Ohio claimed that their algorithm was now better at detecting emotions than people are.

AI systems and devices will soon recognize, interpret, process, and simulate human emotions. A combination of facial analysis, voice pattern analysis, and deep learning can already decode human emotions for market research and political polling purposes. With companies like Affectiva, Beyond Verbal, and Sensay providing plug-and-play sentiment analysis software, the affective computing market is estimated to grow to $41 billion by 2022, as firms like Amazon, Google, Facebook, and Apple race to decode their users' emotions.

Emotional inputs will create a shift from data-driven IQ-heavy interactions to deep EQ-guided experiences, giving brands the opportunity to connect to customers on a much deeper, more personal level. But reading people's emotions is a delicate business. Emotions are highly personal, and users will have concerns about privacy invasion and manipulation. Before companies dive in, leaders should consider questions like:

1. What are you offering? Does your value proposition naturally lend itself to the involvement of emotions? And can you credibly justify the inclusion of emotional clues for the betterment of the user experience?
2. What are your customers' emotional intentions when interacting with your brand? What is the nature of the interaction?
3. Has the user given you explicit permission to analyze their emotions? Does the user stay in control of their data, and can they revoke their permission at any given time?
4. Is your system smart enough to accurately read and react to a user's emotions?
5. What is the danger in any given situation if the system should fail—danger for the user and/or for the brand?

Keeping those concerns in mind, business leaders should be aware of current applications for emotional

Casual applications use social and emotional cures to secure users' loyalty by acting like new AI BFFs.

AI. These fall roughly into three categories:

1. Systems that use emotional analysis to adjust their response. In this application, the AI service acknowledges emotions and factors them into its decision-making process. However, the service's output is completely emotion-free.

Conversational IVRs (interactive voice response) and chatbots that promise to route customers to the right service flow faster and more accurately when factoring in emotions. For example, when the system detects that a user is angry, they are routed to a different escalation flow or to a human.

AutoEmotive, Affectiva's automotive AI, and Ford are racing to get emotional car software market-ready to detect human emotions such as anger or lack of attention, and then take control over or stop the vehicle, preventing accidents or acts of road rage.

The security sector also dabbles in emotional AI to detect stressed or angry people. The British government, for instance, monitors its citizens' sentiments on certain topics over social media.

In this category, emotions play a part in the machine's decision-making process. However, the machine still reacts like a machine—essentially, as a giant switchboard routing people in the right direction.

2. Systems that provide a targeted emotional analysis for learning purposes. In 2009 Philips teamed up with a Dutch bank to develop the idea of a "rationalizer" bracelet to stop traders from making irrational decisions by monitoring their stress levels, which it measures by monitoring the wearer's pulse. Making traders aware of their heightened emotional states made them pause and think before making impulse decisions.

Brain Power's smart glasses help people with autism better understand emotions and social cues. The wearer of this Google Glass–type device sees and hears special feedback geared to the situation—for example, coaching on facial expressions of emotions, when to look at people, and even feedback on the user's own emotional state.

These targeted emotional analysis systems acknowledge and interpret emotions. The insights are communicated to the user for learning purposes. On a personal level, these targeted applications will act like a Fitbit for the heart and mind, aiding in mindfulness, self-awareness, and ultimately self-improvement, while maintaining a machine–person relationship that keeps the user in charge.

Targeted emotional learning systems are also being tested for group settings, such as by analyzing the emotions of students for teachers, or workers for managers. Scaling to group settings can have an Orwellian feeling: Concerns about privacy, creativity, and individuality have these experiments playing on the edge of ethical acceptance. More important, adequate psychological training for the people in power is required to interpret the emotional results and make adequate adjustments.

3. Systems that mimic and ultimately replace human-to-human interactions. When smart speakers entered the American living room in 2014, we started to get used to hearing computers refer to themselves as "I." Call it a human error or an evolutionary shortcut, but when machines talk, people assume relationships.

Products and services can now use conversational user interfaces (UIs) and the concept of "computers as social actors" to try to alleviate mental health concerns. These applications aim to coach users through crises using behavioral therapy techniques. Ellie helps treat soldiers with PTSD. Karim helps Syrian refugees overcome trauma. Digital assistants are even tasked with helping alleviate loneliness among the elderly.

Casual applications like Microsoft's Xiaoice, Google Assistant, or Amazon's Alexa use social and emotional cues for a less altruistic purpose—their aim is to secure users' loyalty by acting like new AI BFFs. Futurist Richard van Hooijdonk quips: "If a marketer can get you to cry, he can get you to buy."

The discussion around addictive technology is starting to examine the intentions behind voice assistants. What does it mean for users if personal assistants are hooked up to advertisers? In a leaked Facebook memo, for example, the social media company boasted to advertisers that it could detect, and subsequently target, teens' feelings of "worthlessness" and "insecurity," among other emotions.

Judith Masthoff of the University of Aberdeen says, "I would like people to have their own guardian angel that could support them emotionally throughout the day." But

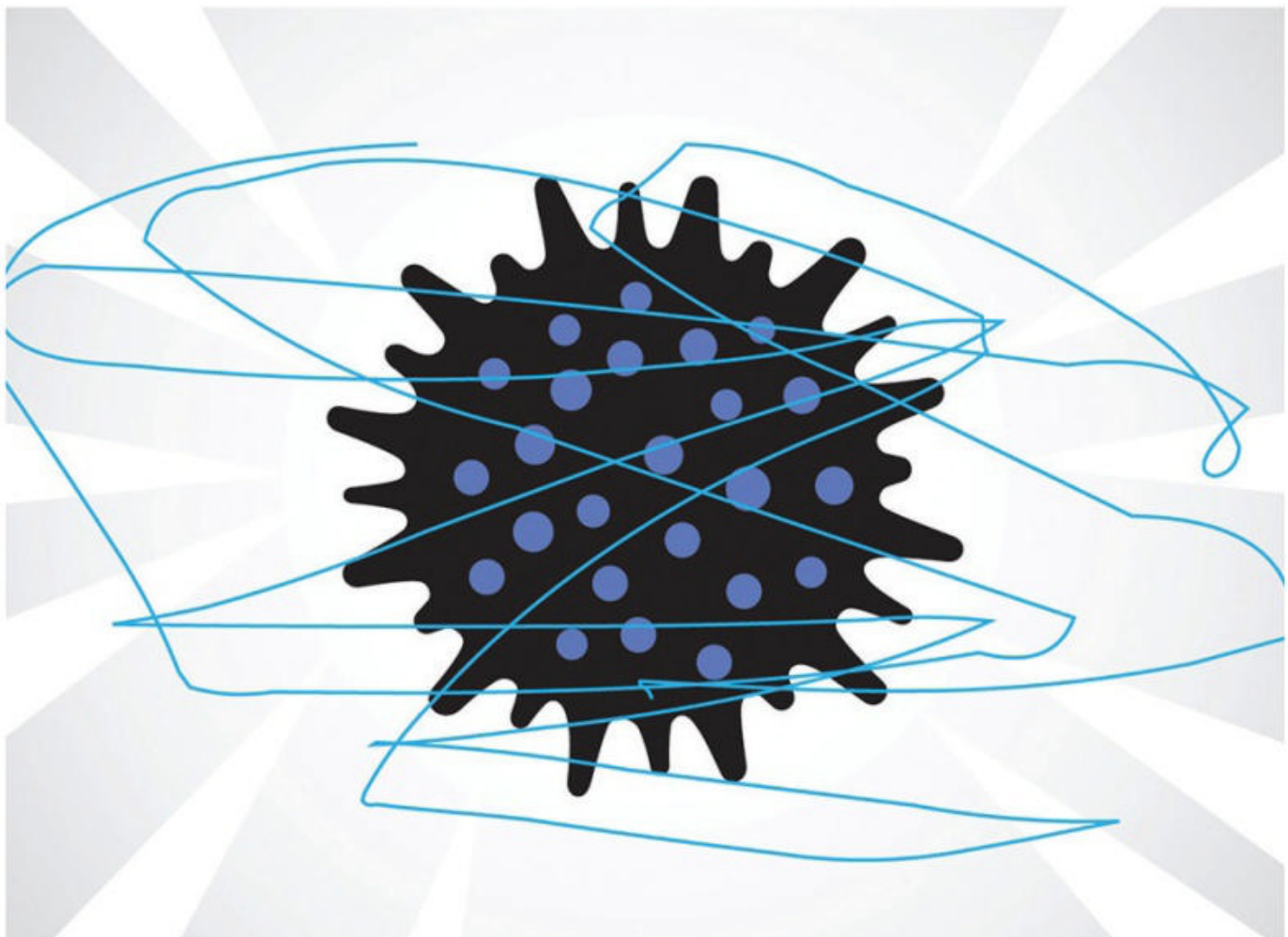
to get to that ideal, a series of (collectively agreed upon) experiments will need to guide designers and brands toward the appropriate level of intimacy, and a series of failures will determine the rules for maintaining trust, privacy, and emotional boundaries.

The biggest hurdle to finding the right balance might not be achieving more-effective forms of emotional AI but finding emotionally intelligent humans to build them.

HBR Reprint H04GCW

Originally published on HBR.org July 31, 2018

***Sophie Kleber** is the executive creative director at Huge, where she creates future-forward user experiences to help transform businesses and influence the way we use technology. From screen-based interfaces, to voice UIs, to sentient computing, her work shapes some of the largest companies in the world, including IKEA, Under Armour, Goldman Sachs, and Warner Brothers. She is a thought leader in the future of design, has written on behalf of TED, and speaks regularly at SXSW.*



HBR STAFF

5. How Hospitals Are Using AI to Battle Covid-19

→ by KELLEY A. WITTBOLD, COLLEEN CARROLL, MARCO IANSITI, HAIPENG (MARK) ZHANG, and ADAM B. LANDMAN

ON MONDAY, MARCH 9, in an effort to address soaring patient demand in Boston, Partners HealthCare went live with a hotline for patients, clinicians, and anyone else with questions and concerns about Covid-19. The goals were to identify and reassure the people who did not need additional care (the vast majority of callers), to offer people with less serious symptoms relevant information and virtual care options, and to direct the smaller number of high-risk and higher-acuity patients to the most appropriate resources, including testing sites, newly created respiratory illness clinics, or in certain cases, emergency departments. As the hotline became overwhelmed, the average wait time peaked at 30 minutes. Many callers gave up before they could speak with the expert team of nurses staffing the hotline. We were missing

While we race to dampen the virus's spread, we can optimize our response mechanisms, digitizing as many steps as possible.

opportunities to facilitate prehospital triage to get the patient to the right care setting at the right time.

The Partners team, led by Lee Schwamm, Haipeng (Mark) Zhang, and Adam Landman, began considering technology options to address the growing need for patient self-triage, including interactive voice response systems and chatbots. We connected with Providence St. Joseph Health system in Seattle, which served some of the country's first Covid-19 patients in early March. In collaboration with Microsoft, Providence built an online screening and triage tool that could rapidly differentiate between those who might really be sick with Covid-19 and those who appeared to be suffering from less threatening ailments. In its first week, Providence's tool served more than 40,000 patients, delivering care at an unprecedented scale.

Our team saw potential for this type of AI-based solution and worked to make a similar tool available to our patient population. The Partners Covid-19 Screener provides a simple, straightforward chat interface, presenting patients with a series of questions based on content from the U.S. Centers for Disease Control and Prevention (CDC) and Partners HealthCare experts. In this way, it too can screen enormous numbers of people and rapidly differentiate between those who might really be sick with Covid-19 and those who are most likely suffering from less threatening ailments. We anticipate this AI bot will alleviate high volumes of patient traffic to the hotline, and extend and stratify the system's care in ways that would have been unimaginable until recently. Development is now underway to facilitate triage of patients with symptoms to the most appropriate care setting, including virtual urgent care, primary care providers, respiratory illness clinics, or emergency departments. Most important, the chatbot can also serve as a near-instantaneous dissemination method for supporting our widely distributed providers, as we have seen the need for frequent clinical triage algorithm updates based on a rapidly changing landscape.

Similarly, at both Brigham and Women's Hospital and Massachusetts General Hospital, physician researchers are exploring the potential use of intelligent robots developed at Boston Dynamics and MIT to deploy in Covid surge clinics and inpatient wards to perform tasks (obtaining vital signs or delivering medication) that would otherwise require human contact in an effort to mitigate disease transmission.

AI Initiatives Are Already Emerging

Several governments and hospital systems around the world have leveraged AI-powered sensors to support triage in sophisticated ways. Chinese technology company Baidu developed a no-contact infrared sensor system to quickly single out individuals with a fever, even in crowds. Beijing's Qinghe railway station is equipped with this system to identify potentially contagious individuals, replacing a cumbersome manual screening process. Similarly, Florida's Tampa General Hospital deployed an AI system in collaboration with Care.ai at its entrances to intercept individuals with potential Covid-19 symptoms from visiting patients. Through cameras positioned at entrances, the technology conducts a facial thermal scan and picks up on other symptoms, including sweat and discoloration, to ward off visitors with fever.

Beyond screening, AI is being used to monitor Covid-19 symptoms, provide decision support for CT scans, and automate hospital operations. Meanwhile, Zhongnan Hospital in China uses an AI-driven CT scan interpreter that identifies Covid-19 when radiologists aren't available. China's Wuhan Wuchang Hospital established a smart field hospital staffed largely by robots. Patients' vital signs are monitored using connected thermometers and bracelet-like devices. Intelligent robots deliver medicine and food to patients, alleviating physician exposure to the virus and easing the workload of health care workers experiencing exhaustion. And in South Korea, the government released an app that allows users to self-report symptoms and alerts them if they leave a "quarantine zone" in order to curb the impact of super-spreaders who would otherwise go on to infect large populations.

Digital Transformation Now

The spread of Covid-19 is stretching operational systems in health care and beyond. We have seen shortages of everything, from masks and gloves to ventilators, and from emergency room capacity and ICU beds to the speed and reliability of internet connectivity. The

reason is both simple and terrifying: Our economy and health care systems are geared to handle linear, incremental demand, whereas the virus grows at an exponential rate. Our national health system cannot keep up with this kind of explosive demand without the rapid and large-scale adoption of digital operating models.

While we race to dampen the virus's spread, we can optimize our response mechanisms, digitizing as many steps as possible. This is because traditional processes that rely on people to function in the critical path of signal processing are constrained by the rate at which we can train, organize, and deploy human labor. Moreover, traditional processes deliver decreasing returns as they scale. On the other hand, digital systems can be scaled up without such constraints, at virtually infinite rates. The only theoretical bottlenecks are computing power and storage capacity—and we have plenty of both. Digital systems can keep pace with exponential growth.

Importantly, AI for health care must be balanced by the appropriate level of human clinical expertise for final decision-making to ensure we are delivering high-quality, safe care. In many cases, human clinical reasoning and decision making cannot be easily replaced by AI; instead, AI is a decision aid that helps humans improve effectiveness and efficiency.

Digital transformation in health care has been lagging in other industries. Our response to Covid today has accelerated the adoption and scaling of virtual and AI tools. From the AI bots deployed by Providence and Partners HealthCare to the Smart Field Hospital in Wuhan, rapid digital transformation is being employed to tackle the exponentially growing Covid threat. We hope and anticipate that after Covid-19 settles, we will have transformed the way we deliver health care in the future.

HBR Reprint H05IWJ

Originally published on HBR.org April 3, 2020

Kelley A. Wittbold *is an emergency medicine attending physician at Massachusetts General Hospital and an administrative fellow specializing in digital health, emergency telemedicine, and emergency operations management.* ***Colleen Carroll*** *is a principal at Keystone Strategy. Her work focuses on digital health strategies and health care economics for biotechnology firms, pharmaceutical companies, and health care systems.* ***Marco Iansiti*** *is the David Sarnoff Professor of Business Administration at Harvard Business School, where he heads the Digital Initiative and is faculty cochair of the Laboratory for Innovation Science. He has advised many companies in the technology sector, including Microsoft, Facebook, and Amazon. He is a coauthor (with Karim Lakhani) of* Competing in the Age of AI *(Harvard Business Review Press, 2020).* ***Haipeng (Mark) Zhang*** *is the associate chief medical information officer at Brigham Hospital and a palliative care consultant at Dana-Farber Cancer Institute. He is also the medical director of the Brigham and Women's Digital Innovation Hub and focuses on the application and creation of digital tools for clinicians, staff, and patients at Brigham Health and throughout the Partners HealthCare enterprise.* ***Adam B. Landman*** *is an emergency medicine and clinical informatics attending physician at Brigham and Women's Hospital, where he also serves as vice president and chief information and digital officer. He is responsible for developing system-wide strategic IT initiatives, with the goal of evolving the next generation of information systems across the Brigham Health enterprise.*

ORIGINALLY PUBLISHED JANUARY–FEBRUARY 2018

Artificial Intelligence for the Real World

Don't start with moonshots.

→ by THOMAS H. DAVENPORT and RAJEEV RONANKI

IN 2013, THE MD ANDERSON Cancer Center launched a "moonshot" project: diagnose and recommend treatment plans for certain forms of cancer using IBM's Watson cognitive system. But in 2017, the project was put on hold after costs topped $62 million—and the system had yet to be used on patients. At the same time, the cancer center's IT group was experimenting with using cognitive technologies to do much less ambitious jobs, such as making hotel and restaurant recommendations for patients' families, determining which patients needed help paying bills, and addressing staff IT

Illustration by **ENISAURUS**

Broadly speaking, AI can support three important business needs: automating business processes, gaining insight through data analysis, and engaging with customers and employees.

problems. The results of these projects have been much more promising: The new systems have contributed to increased patient satisfaction, improved financial performance, and a decline in time spent on tedious data entry by the hospital's care managers. Despite the setback on the moonshot, MD Anderson remains committed to using cognitive technology—that is, next-generation artificial intelligence—to enhance cancer treatment, and is currently developing a variety of new projects at its center of competency for cognitive computing.

The contrast between the two approaches is relevant to anyone planning AI initiatives. Our survey of 250 executives who are familiar with their companies' use of cognitive technology shows that three-quarters of them believe that AI will substantially transform their companies within three years. However, our study of 152 projects in almost as many companies also reveals that highly ambitious moonshots are less likely to be successful than "low-hanging fruit" projects that enhance business processes. This shouldn't be surprising—such has been the case with the great majority of new technologies that companies have adopted in the past. But the hype surrounding artificial intelligence has been especially powerful, and some organizations have been seduced by it.

In this article, we'll look at the various categories of AI being employed and provide a framework for how companies should begin to build up their cognitive capabilities in the next several years to achieve their business objectives.

Three Types of AI

It is useful for companies to look at AI through the lens of business capabilities rather than technologies. Broadly speaking, AI can support three important business needs: automating business processes, gaining insight through data analysis, and engaging with customers and employees. (See the exhibit "Cognitive Projects by Type.")

Process automation. Of the 152 projects we studied, the most common type was the automation of digital and physical tasks—typically back-office administrative and financial activities—using robotic process automation technologies. RPA is more advanced than earlier business-process automation tools, because the "robots" (that is, code on a server) act like a human inputting and consuming information from multiple IT systems. Tasks include:

- transferring data from email and call center systems into systems of record—for example, updating customer files with address changes or service additions;
- replacing lost credit or ATM cards, reaching into multiple systems to update records and handle customer communications;
- reconciling failures to charge for services across billing systems by extracting information from multiple document types; and
- "reading" legal and contractual documents to extract provisions using natural language processing.

RPA is the least expensive and easiest to implement of the cognitive technologies we'll discuss here, and typically brings a quick and high return on investment. (It's also the least "smart" in the sense that these applications aren't programmed to learn and improve, though developers are slowly adding more intelligence and learning capability.) It is particularly well suited to working across multiple back-end systems.

At NASA, cost pressures led the agency to launch four RPA pilots in accounts payable and receivable, IT spending, and human resources—all managed by a shared services center. The four projects worked well—in the HR application, for example, 86% of transactions were completed without human intervention—and are being rolled out across the organization. NASA is now implementing more RPA bots, some with higher levels of intelligence. As Jim Walker, project leader for the shared services organization notes, "So far it's not rocket science."

One might imagine that robotic process automation would quickly put people out of work. But across the 71 RPA projects we reviewed (47% of the total), replacing administrative employees was neither the primary objective nor a common outcome. Only a few projects led to reductions in head count, and in most cases, the tasks in question had already been shifted to outsourced workers. As technology improves, robotic automation projects are likely to lead to some job losses in the future, particularly in the offshore business-process outsourcing industry. If you can outsource a task, you can probably automate it.

Cognitive insight. The second most common type of project in our study (38% of the total) used algorithms to detect patterns in vast volumes of data and interpret their meaning. Think of it as "analytics on steroids." These machine-learning applications are being used to:

- predict what a particular customer is likely to buy;
- identify credit fraud in real time and detect insurance claims fraud;
- analyze warranty data to identify safety or quality problems in automobiles and other manufactured products;
- automate personalized targeting of digital ads; and
- provide insurers with more-accurate and detailed actuarial modeling.

Cognitive insights provided by machine learning differ from those available from traditional analytics in three ways: They are usually much more data-intensive and detailed, the models typically are trained on some part of the data set, and the models get better—that is, their ability to use new data to make predictions or put things into categories improves over time.

Versions of machine learning (deep learning, in particular, which attempts to mimic the activity in the human brain in order to recognize patterns) can perform feats such as recognizing images and speech. Machine learning can also make available new data for better analytics. While the activity of data curation has historically been quite labor-intensive, now machine learning can identify probabilistic matches—data that is likely to be associated with the same person or company but that appears in slightly different formats—across databases. GE has used this technology to integrate supplier data and has saved $80 million in its first year by eliminating redundancies and negotiating contracts that were previously managed at the business unit level. Similarly, a large bank used this technology to extract data on terms from supplier contracts and match it with invoice numbers, identifying tens of millions of dollars in products and services not supplied. Deloitte's audit practice is using cognitive insight to extract terms from contracts, which enables an audit to address a much higher proportion of documents, often 100%, without human auditors' having to painstakingly read through them.

Cognitive insight applications are typically used to improve performance on jobs only machines can do—tasks such as programmatic ad buying that involve such high-speed data crunching and automation that they've long been beyond human ability—so they're not generally a threat to human jobs.

Cognitive engagement. Projects that engage employees and customers using natural language processing chatbots, intelligent agents, and machine learning were the least common type in our study (accounting for 16% of the total). This category includes:

- intelligent agents that offer 24/7 customer service, addressing a broad and growing array of issues from password requests to technical support questions—all in the customer's natural language;
- internal sites for answering employee questions on topics including IT, employee benefits, and HR policy;

Idea in Brief

THE PROBLEM
Cognitive technologies are increasingly being used to solve business problems, but many of the most ambitious AI projects encounter setbacks or fail.

THE APPROACH
Companies should take an incremental rather than a transformative approach and focus on augmenting rather than replacing human capabilities.

THE PROCESS
To get the most out of AI, firms must understand which technologies perform what types of tasks, create a prioritized portfolio of projects based on business needs, and develop plans to scale up across the company.

- product and service recommendation systems for retailers that increase personalization, engagement, and sales—typically including rich language or images; and
- health treatment recommendation systems that help providers create customized care plans that take into account individual patients' health status and previous treatments.

The companies in our study tended to use cognitive engagement technologies more to interact with employees than with customers. That may change as firms become more comfortable turning customer interactions over to machines. Vanguard, for example, is piloting an intelligent agent that helps its customer service staff answer frequently asked questions. The plan is to eventually allow customers to engage with the cognitive agent directly, rather than with the human customer-service agents. SEBank, in Sweden, and the medical technology giant Becton, Dickinson, in the United States, are using the lifelike intelligent-agent avatar Amelia to serve as an internal employee help desk for IT support. SEBank has recently made Amelia available to customers on a limited basis in order to test its performance and customer response.

Companies tend to take a conservative approach to customer-facing cognitive engagement technologies largely because of their immaturity. Facebook, for example, found that its Messenger chatbots couldn't answer 70% of customer requests without human intervention. As a result, Facebook and several other firms are restricting bot-based interfaces to certain topic domains or conversation types.

Our research suggests that cognitive engagement apps are not currently threatening customer service or sales rep jobs. In most of the projects we studied, the goal was not to reduce head count but to handle growing numbers of employee and customer interactions without adding staff. Some organizations were planning to hand over routine communications to machines, while transitioning customer-support personnel to more-complex activities such as handling customer issues that escalate, conducting extended unstructured dialogues, or reaching out to customers before they call in with problems.

As companies become more familiar with cognitive tools, they are experimenting with projects that combine elements from all three categories to reap the benefits of AI. An Italian insurer, for example, developed a "cognitive help desk" within its IT organization. The system engages with employees using deep-learning technology (part of the cognitive insights category) to search frequently asked questions and answers, previously resolved cases, and documentation to come up with solutions to employees' problems. It uses a smart-routing capability (business process automation) to forward the most complex problems to human representatives, and it uses natural language processing to support user requests in Italian.

Despite their rapidly expanding experience with cognitive tools, however, companies face significant obstacles in development and implementation. On the basis of our research, we've developed a four-step framework for integrating AI technologies that can help companies achieve their objectives, whether the projects are moon shoots or business-process enhancements.

1 Understanding the Technologies

Before embarking on an AI initiative, companies must understand which technologies perform what types of tasks, and the strengths and limitations of each. Rule-based expert systems and robotic process automation, for example, are transparent in how they do their work, but neither is capable of learning and improving. Deep learning, on the other hand, is great at learning from large volumes of labeled data, but it's almost impossible to understand how it creates the models it does. This "black box" issue can be problematic in highly regulated industries such as financial services, in which regulators insist on knowing why decisions are made in a certain way.

We encountered several organizations that wasted time and money pursuing the wrong technology for the job at hand. But if they're armed with a good understanding of the different technologies, companies are better positioned to determine which might best address specific needs, which vendors to work with, and how quickly a system can be implemented. Acquiring this understanding requires ongoing research and education, usually within IT or an innovation group.

Cognitive Projects by Type

We studied 152 cognitive technology projects and found that they fell into three categories.

Robotics & Cognitive Automation	Cognitive Insight	Cognitive Engagement
71	57	24

In particular, companies will need to leverage the capabilities of key employees, such as data scientists, who have the statistical and big-data skills necessary to learn the nuts and bolts of these technologies. A main success factor is your people's willingness to learn. Some will leap at the opportunity, while others will want to stick with tools they're familiar with. Strive to have a high percentage of the former.

If you don't have data science or analytics capabilities in-house, you'll probably have to build an ecosystem of external service providers in the near term. If you expect to be implementing longer-term AI projects, you will want to recruit expert in-house talent. Either way, having the right capabilities is essential to progress.

Given the scarcity of cognitive technology talent, most organizations should establish a pool of resources—perhaps in a centralized function such as IT or strategy—and make experts available to high-priority projects throughout the organization. As needs and talent proliferate, it may make sense to dedicate groups to particular business functions or units, but even then a central coordinating function can be useful in managing projects and careers.

2 Creating a Portfolio of Projects

The next step in launching an AI program is to systematically evaluate needs and capabilities and then develop a prioritized portfolio of projects. In the companies we studied, this was usually done in workshops or through small consulting engagements. We recommend that companies conduct assessments in three broad areas.

Identifying the opportunities. The first assessment determines which areas of the business could benefit most from cognitive applications. Typically, they are parts of the company where "knowledge"—insight derived from data analysis or a collection of texts—is at a premium but for some reason is not available.

- **Bottlenecks.** In some cases, the lack of cognitive insights is caused by a bottleneck in the flow of information; knowledge exists in the organization, but it is not optimally distributed. That's often the case in health care, for example, where knowledge tends to be siloed within practices, departments, or academic medical centers.
- **Scaling challenges.** In other cases, knowledge exists, but the process for using it takes too long or is expensive to scale. Such is often the case with knowledge developed by financial advisers. That's why many investment and wealth management firms now offer AI-supported "robo-advice" capabilities that provide clients with cost-effective guidance for routine financial issues.

In the pharmaceutical industry, Pfizer is tackling the scaling problem by using IBM's Watson to accelerate the laborious process of drug-discovery research in immuno-oncology, an emerging approach to cancer treatment that uses the body's immune system to help fight cancer. Immuno-oncology drugs can take up to 12 years to bring to market. By combining a sweeping literature review with Pfizer's own data, such as lab reports, Watson is helping researchers to surface relationships and find hidden patterns that should speed the identification of new drug targets, combination therapies for study, and patient selection strategies for this new class of drugs.

- **Inadequate firepower.** Finally, a company may collect more data than its existing human or computer firepower can adequately analyze and apply. For example, a company may have massive amounts of data on consumers' digital behavior but lack insight about what it means or how it can be strategically applied. To address this, companies are using machine learning to support tasks such as programmatic buying of personalized digital ads or, in the case of Cisco Systems and IBM, to create tens of thousands of "propensity models" for determining which customers are likely to buy which products.

Determining the use cases. The second area of assessment evaluates the use cases in which cognitive applications would generate substantial value and contribute to business success. Start by asking key questions such as: How critical to your overall strategy is addressing the targeted problem? How difficult would it be to implement the proposed AI solution—both technically and organizationally? Would the benefits from launching the application be worth the effort? Next, prioritize the use cases according to which offer the most short- and long-term value, and which might ultimately be integrated into a broader platform or suite of cognitive capabilities to create competitive advantage.

If your firm plans to launch several pilots, consider creating a cognitive center of excellence or similar structure to manage them.

The Business Benefits of AI

We surveyed 250 executives who were familiar with their companies' use of cognitive technologies to learn about their goals for AI initiatives. More than half said their primary goal was to make existing products better. Reducing head count was mentioned by only 22%.

Percentage of executives who cite the following as benefits of AI

Source: Deloitte 2017

Selecting the technology. The third area to assess examines whether the AI tools being considered for each use case are truly up to the task. Chatbots and intelligent agents, for example, may frustrate some companies because most of them can't yet match human problem-solving beyond simple scripted cases (though they are improving rapidly). Other technologies, like robotic process automation that can streamline simple processes such as invoicing, may in fact slow down more-complex production systems. And while deep learning visual recognition systems can recognize images in photos and videos, they require lots of labeled data and may be unable to make sense of a complex visual field.

In time, cognitive technologies will transform how companies do business. Today, however, it's wiser to take incremental steps with the currently available technology while planning for transformational change in the not-too-distant future. You may ultimately want to turn customer interactions over to bots, for example, but for now it's probably more feasible—and sensible—to automate your internal IT help desk as a step toward the ultimate goal.

3 Launching Pilots

Because the gap between current and desired AI capabilities is not always obvious, companies should create pilot projects for cognitive applications before rolling them out across the entire enterprise.

Proof-of-concept pilots are particularly suited to initiatives that have high potential business value or allow the organization to test different technologies at the same time. Take special care to avoid "injections" of projects by senior executives who have been influenced by technology vendors. Just because executives and boards of directors may feel pressure to "do something cognitive" doesn't mean you should bypass the rigorous piloting process. Injected projects often fail, which can significantly set back the organization's AI program.

If your firm plans to launch several pilots, consider creating a cognitive center of excellence or similar structure to manage them. This approach helps build the needed technology skills and capabilities within the organization, while also helping to move small pilots into broader applications that will have a greater impact. Pfizer has more than 60 projects across the company that employ some form of cognitive technology; many are pilots, and some are now in production.

At Becton, Dickinson, a "global automation" function within the IT organization oversees a number of cognitive technology pilots that use intelligent digital agents and RPA (some work is done in partnership with the company's Global Shared Services organization). The global automation group uses end-to-end process maps to guide implementation and identify automation opportunities. The group also uses graphical "heat maps" that indicate the organizational activities most amenable to AI interventions. The company has successfully implemented intelligent agents in IT support processes, but as yet is not ready to support large-scale enterprise processes, like order-to-cash.

The health insurer Anthem has developed a similar centralized AI function that it calls the Cognitive Capability Office.

Business-process redesign. As cognitive technology projects are developed, think through how workflows might be redesigned, focusing specifically on the division of labor between humans and the AI. In some cognitive projects, 80% of decisions will be made by machines and 20% will be made by humans; others will have the opposite ratio. Systematic redesign of workflows is necessary to ensure that humans and machines augment each other's strengths and compensate for weaknesses.

The investment firm Vanguard, for example, has a new "Personal Advisor Services" (PAS) offering, which combines automated investment advice with guidance from human advisers. In the new system, cognitive technology is used to perform many of the traditional tasks of investment advising, including constructing a customized portfolio, rebalancing portfolios over time, tax-loss harvesting, and tax-efficient investment selection. Vanguard's human advisers serve as "investing coaches," tasked with answering investor questions, encouraging healthy financial behaviors, and being, in Vanguard's words, "emotional circuit breakers" to keep investors on plan. Advisers are encouraged to learn about behavioral finance to perform these roles effectively. The PAS approach has quickly gathered more than $80 billion in assets under management, costs are lower than those for purely human-based advising, and customer satisfaction is high. (See the exhibit "One Company's Division of Labor.")

Vanguard understood the importance of work redesign when implementing PAS, but many companies simply "pave the cow path" by automating existing work processes, particularly when using RPA technology. By automating established workflows, companies can quickly implement projects and achieve ROI—but they forgo the opportunity to take full advantage of AI capabilities and substantively improve the process.

Cognitive work redesign efforts often benefit from applying design-thinking principles: understanding customer or end-user needs, involving employees whose work will be restructured, treating designs as experimental "first drafts," considering multiple alternatives, and explicitly considering cognitive technology capabilities in the design process. Most cognitive projects are also suited to iterative, agile approaches to development.

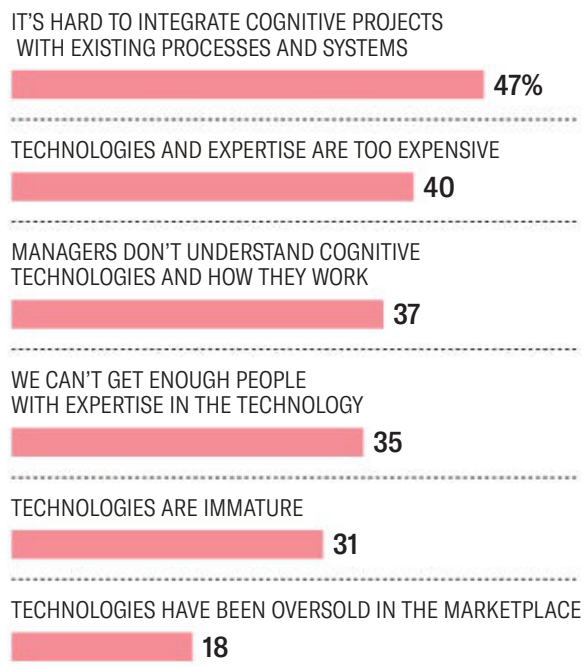
4 Scaling Up

Many organizations have successfully launched cognitive pilots, but they haven't had as much success rolling them out organization-wide. To achieve their goals, companies need detailed plans for scaling up, which requires collaboration between technology experts and owners of the business process being automated. Because cognitive technologies typically support individual tasks rather than entire processes, scale-up almost always requires integration with existing systems and processes. Indeed, in our survey, executives reported that

The Challenges of AI

Executives in our survey identified several factors that can stall or derail AI initiatives, ranging from integration issues to scarcity of talent.

Percentage who cite the following as obstacles



Source: Deloitte 2017

One Company's Division of Labor

Vanguard, the investment services firm, uses cognitive technology to provide customers with investment advice at a lower cost. Its Personal Advisor Services system automates many traditional tasks of investment advising, while human advisers take on higher-value activities. Here's how Vanguard redesigned its work processes to get the most from the new system.

Cognitive technology

- Generates a financial plan
- Provides goals-based forecasting in real time
- Rebalances portfolio to target mix
- Minimizes taxes
- Tracks aggregated assets in one place
- Engages clients virtually

Adviser

- Understands investment goals
- Customizes an implementation plan
- Provides investment analysis and retirement planning
- Develops retirement income and Social Security drawdown strategies
- Serves as a behavioral coach
- Monitors spending to encourage accountability
- Offers ongoing wealth and financial-planning support
- Addresses estate-planning considerations

Source: Vanguard group

such integration was the greatest challenge they faced in AI initiatives.

Companies should begin the scaling-up process by considering whether the required integration is even possible or feasible. If the application depends on special technology that is difficult to source, for example, that will limit scale-up. Make sure your business process owners discuss scaling considerations with the IT organization before or during the pilot phase: An end run around IT is unlikely to be successful, even for relatively simple technologies like RPA.

The health insurer Anthem, for example, is taking on the development of cognitive technologies as part of a major modernization of its existing systems. Rather than bolting new cognitive apps onto legacy technology, Anthem is using a holistic approach that maximizes the value being generated by the cognitive applications, reduces the overall cost of development and integration, and creates a halo effect on legacy systems. The company is also redesigning processes at the same time to, as CIO Tom Miller puts it, "use cognitive to move us to the next level."

In scaling up, companies may face substantial change-management challenges. At one U.S. apparel retail chain, for example, the pilot project at a small subset of stores used machine learning for online product recommendations, predictions for optimal inventory and rapid replenishment models, and—most difficult of all—merchandising. Buyers, used to ordering product on the basis of their intuition, felt threatened and made comments like "If you're going to trust this, what do you need me for?" After the pilot, the buyers went as a group to the chief merchandising officer and requested that the program be killed. The executive pointed out that the results were positive and warranted expanding the project. He assured the buyers that, freed of certain merchandising tasks, they could take on more high-value work that humans can still do better than machines, such as understanding younger customers' desires and determining apparel manufacturers' future plans. At the same time, he acknowledged that the merchandisers needed to be educated about a new way of working.

If scale-up is to achieve the desired results, firms must also focus on improving productivity. Many, for example, plan to grow their way into productivity—adding customers and transactions without adding staff. Companies that cite head count reduction as the primary justification for the AI investment should ideally plan to realize that goal over time through attrition or from the elimination of outsourcing.

The Future Cognitive Company

Our survey and interviews suggest that managers experienced with cognitive technology are bullish on its prospects. Although the early successes are relatively modest, we anticipate that these technologies will eventually transform work. We believe that companies that are adopting AI in moderation now—and have aggressive implementation plans for the future—will find themselves as well positioned to reap benefits as those that embraced analytics early on.

With the right planning and development, cognitive technology could usher in a golden age of productivity, work satisfaction, and prosperity.

Through the application of AI, information-intensive domains such as marketing, health care, financial services, education, and professional services could become simultaneously more valuable and less expensive to society. Business drudgery in every industry and function—overseeing routine transactions, repeatedly answering the same questions, and extracting data from endless documents—could become the province of machines, freeing up human workers to be more productive and creative. Cognitive technologies are also a catalyst for making other data-intensive technologies succeed, including autonomous vehicles, the Internet of Things, and mobile and multichannel consumer technologies.

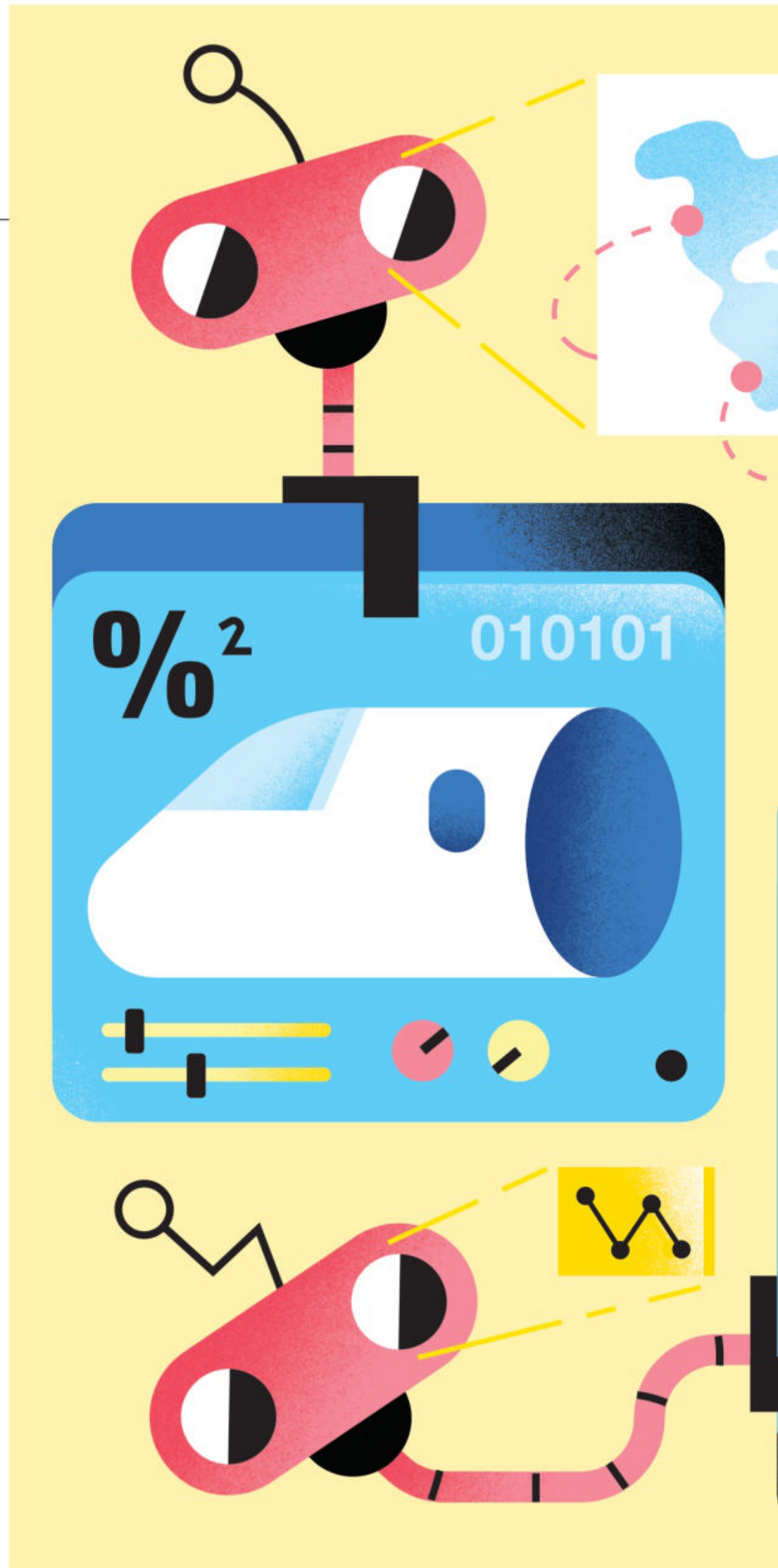
The great fear about cognitive technologies is that they will put masses of people out of work. Of course, some job loss is likely as smart machines take over certain tasks traditionally done by humans. However, we believe that most workers have little to fear at this point. Cognitive systems perform tasks, not entire jobs. The human job losses we've seen were primarily due to attrition of workers who were not replaced or through automation of outsourced work. Most cognitive tasks currently being performed augment human activity, perform a narrow task within a much broader job, or do work that wasn't done by humans in the first place, such as big-data analytics.

Most managers with whom we discuss the issue of job loss are committed to an augmentation strategy—that is, integrating human and machine work, rather than replacing humans entirely. In our survey, only 22% of executives indicated that they considered reducing head count as a primary benefit of AI.

We believe that every large company should be exploring cognitive technologies. There will be some bumps in the road, and there is no room for complacency on issues of workforce displacement and the ethics of smart machines. But with the right planning and development, cognitive technology could usher in a golden age of productivity, work satisfaction, and prosperity.

HBR Reprint R1801H

Thomas H. Davenport *is the President's Distinguished Professor of Information Technology and Management at Babson College, a research fellow at the MIT Initiative on the Digital Economy, and a senior adviser at Deloitte Analytics. He is the author of over a dozen management books, most recently* Only Humans Need Apply: Winners and Losers in the Age of Smart Machines *(Harper Business, 2016), with coauthor Julia Kirby, and* The AI Advantage: How to Put the Artificial Intelligence Revolution to Work *(MIT Press, 2018).* ***Rajeev Ronanki*** *is a Senior Vice President and Chief Digital Officer at Anthem, Inc. He leads Anthem's transformation into a digital platform for health. Rajeev also is the author of* You And AI: A Citizen's Guide to AI, Blockchain, and Puzzling Together the Future of Healthcare.



ORIGINALLY PUBLISHED MAY–JUNE 2021

Getting AI to Scale

Don't try to change everything at once, but do begin with something important.

→ by TIM FOUNTAINE, BRIAN MCCARTHY, and TAMIM SALEH

MOST CEOS RECOGNIZE that artificial intelligence has the potential to completely change how organizations work. They can envision a future in which, for example, retailers deliver individualized products before customers even request them—perhaps on the very same day those products are made. That scenario may sound like science fiction, but the AI that makes it possible already exists.

What's getting in the way of that future is that companies haven't figured out how to change themselves to meet it. To be fair, most have been working hard to incorporate digital technologies, in some instances genuinely transforming the way they serve their customers and manufacture their offerings.

To capture the full promise of AI, however, companies must reimagine their business models and the way work

Illustration by **ENISAURUS**

0001
010101
101
010101

gets done. They can't just plug AI into an existing process to automate it or add insights. And while AI can be employed locally across functions in a laundry list of specific applications (known as *use cases*), that approach won't drive consequential change in a company's operations or bottom line. It also makes it much harder and more costly to get AI to scale, because each far-flung team must reinvent the wheel with respect to stakeholder buy-in, training, change management, data, technology, and more.

But that doesn't mean companies should try to overhaul the whole organization with AI all at once. That would almost certainly end in failure. A complete makeover is an enormously complicated process involving too many moving parts, stakeholders, and projects to achieve meaningful impact quickly.

The right approach, we've found, is to identify a crucial slice of the business and rethink it completely. Introducing changes throughout an entire core process, journey, or function—what we call a *domain*—will lead to a major improvement in performance that isolated local applications simply cannot match. It also will enable each AI initiative to build off the previous one by, for example, reusing data or advancing capabilities for a common set of stakeholders. We've seen this approach trigger an organic cycle of change within domains and, ultimately, build momentum for the use of AI throughout the larger organization as business leaders and employees see it work. Moreover, this approach promotes a mindset of continuous improvement in the workforce, which is crucial because AI technology is advancing rapidly, requiring organizations to think of AI transformations as ongoing rather than one-time efforts.

Ultimately, the companies that can't take full advantage of AI will be sidelined by those that can—as we already see happening in several industries, like auto manufacturing and financial services. The good news is that over the past year many companies (even firms with limited analytics capabilities) have begun developing the skills required to capture AI opportunities, as the Covid-19 crisis forced them to alter the way they did business almost overnight. Now the challenge will be applying those skills to pull off larger initiatives.

In the following pages, we'll draw on our experience working with hundreds of clients, including some of the world's largest organizations, to describe what companies need to do to get AI to scale.

STEP 1

Set the Strategy

It can be challenging to get the scope of AI initiatives just right. We advise CEOs to target areas of the business where AI will make a big difference in a reasonable period of time, it's relatively easy to find a sponsor, get stakeholders to buy in, and put together a team; and there are multiple interconnected activities and opportunities to reuse data and technology assets. (To find out if you haven't scoped your initiatives correctly, see the sidebar "Signs You're Thinking About AI Too Broadly or Too Narrowly.")

Potential impact. The chosen domains should be large enough to significantly improve either the company's bottom line or customer or employee experiences. One airline we advise determined that it had 10 main business domains fitting that description: cargo, crew, revenue management, e-commerce, customer service, airports, maintenance, network planning, operations, and talent. But it started with cargo, where it had identified a portfolio of AI initiatives that could be completed in about 18 weeks. The first would deliver some $30 million in additional profit by enabling more accurate forecasting of cargo volumes and weight and increasing the use of shipping capacity.

In another case a telecom provider chose to redesign its process for managing customer value (which spans all the ways a company interacts with its customers), using AI to understand and address each customer's unique needs. That work quickly reduced the time it took to execute marketing campaigns by 75% and enabled the company to lower customer churn by three percentage points. The company expects those improvements to add $70 million to its bottom line by the end of 2021.

Interconnected activities. Promising domains encompass a clear-cut set of business activities whose recalibration can solve systemic problems like chronic inefficiencies (such as lengthy loan approval times), high variability (rapidly fluctuating consumer demand), and routinely missed opportunities (difficulties getting products to customers). In many cases AI solutions may address the root causes of these problems, partly through the insights delivered

There will be a start-up investment for the first model or two created, but over time new projects can build off past ones, dramatically reducing development time and cost.

and partly through organizational improvements.

The airline identified six closely intertwined cargo activities: negotiating rates, allocating space, booking reservations, documenting shipments, managing ground operations and delivery, and billing. Customer satisfaction and pricing were both dependent on factors such as the availability of space on short notice, the ability to track shipments in real time, and the speed of delivery. When the six activities were reconfigured so that they could feed data into an AI-supported platform, the company was able to significantly reduce systemic waste while greatly improving the customer experience—bolstering its margins and its reputation at the same time.

Sponsor and team. In a promising domain you can readily identify the following:

- an internal business champion responsible for the entire value chain involved (at the airline, it was the vice president of cargo)
- dedicated senior business staff (at the airline this included the senior director of cargo and two of his direct reports) who can fill the roles of "product owner" (the person responsible for solution delivery), translator (who bridges the analytics and business realms), and change lead (responsible for change management efforts)
- a team of AI practitioners, such as data science and engineering experts, designers, business analysts, and a scrum master (these practitioners may also be drawn from a central team in the organization)
- a cluster of frontline users or knowledge workers responsible for day-to-day activities (at the airline, they included 250 sales and reservation agents across the Americas, Asia Pacific, and Europe)

Drafting employees from across the domain life cycle (regardless of where they formerly sat within the organization) and giving them accountability for the work builds engagement with an initiative and creates excitement and momentum. Those factors are crucial to getting employees to think beyond business as usual in devising solutions and help the project clear inevitable unexpected hurdles.

Reusable technology and data. It's also important to select domains where the data and technology components necessary to run the AI models can overlap. It's much easier when teams don't have to start from scratch every time and can reuse data or snippets of code that have already been prepared for AI. There will likely be a start-up investment for the first model or two created within a domain, but over time new projects can build off past ones, dramatically reducing development time and cost. The resources we're referring to here often include, on the data side, common libraries and metadata definitions, and on the technology side, machine learning scripts, application programming interfaces (APIs) that extract data from legacy systems, and data visualization capabilities.

Executive teams typically will identify about eight to 10 domains where AI can transform their business. Once they do, we recommend that they winnow

Idea in Brief

THE PROBLEM
Most companies aren't setting themselves up to realize the full potential of AI. That's because they focus on applying it in discrete use cases, which delivers only incremental change and requires much more effort to scale up.

THE SOLUTION
Organizations are most successful when they reimagine a core business process, journey, or function enabled by AI end to end. That allows each AI effort to build off the previous one, triggering an organic cycle of change.

HOW TO MAKE IT HAPPEN
Leaders must help their organizations identify business domains where AI can make a big difference and target one or two for a complete overhaul. That will involve deploying new technology, redesigning operational processes, changing how people work together, and even fundamentally rethinking business models.

Teams should consider the potential impact that AI initiatives will have on upstream and downstream processes and implement measures to address it.

Signs You're Thinking About AI Too Broadly or Too Narrowly

TOO BROADLY	TOO NARROWLY
The work identified in one domain can't be completed within three or four waves of work over 12 to 15 months.	You're solving a niche challenge while leaving the root causes of problems untouched or not taking into account interrelated processes.
There are more than a dozen leaders with different goals who get to say what should happen next and there's no clear business owner with accountability.	The business leader in the target area doesn't feel ownership because the project won't move the needle, and you haven't involved leaders from across a specific value chain.
You need to redesign the whole data and tech architecture of the company to get any value.	You've created a solution that doesn't integrate with other upstream and downstream processes.

the list down to one or two on the basis of feasibility and business value.

At the airline the CEO and his direct reports had held a series of strategy sessions over 12 weeks. They discussed how companies across different industries were innovating with AI, developed a vision for using AI to achieve a double-digit increase in operating profit within 15 months, prioritized which domains to start with, and committed the resources required to move forward. The executives each asked experts within their individual domains to identify what their areas could do differently to reach the profit goal and to assess the potential value and feasibility of their recommendations. In the cargo domain three senior business leaders, along with IT and finance staff, sketched out the opportunity to better fill available cargo space on planes, the expected returns for doing so, and the practicality of accomplishing this in terms of data availability, technology, talent, and so on.

STEP 2

Structure the Team

The team responsible for AI initiatives within each domain should contain all the people necessary—from business, digital, analytics, and IT functions—to design, build, and support the new ways of working. To a great extent, once domain teams know their objective and are resourced, they will organize their work on their own, using agile practices. The role of management, beyond creating the teams, will be to ensure that any employees moved onto them from other parts of the business are fully integrated and to remove any organizational barriers that might impede teams' success.

In many cases we've studied, most of the team members needed were already working in the target domain, and leaders simply had to shift them onto the project and then bring in the necessary technical talent from other areas of the company. At the airline, sales, customer service, operations, and finance employees all were involved in the cargo domain transformation, and most of them had reported to the business function from the outset. AI experts, such as data scientists and data engineers, were assigned to the team from the company's AI center of excellence for the duration of the work and reported directly to the senior director in the cargo division, who was the product owner for the new AI.

In some cases companies will have to explicitly reassign people in other, nontechnical roles from various parts of the organization to the team. Consider an energy utility retailer that also sought to use AI to revamp customer value management, including which customers were targeted, which offers were sent to them and through which channels, and how new ideas were tested. The company had to formally move previously siloed marketing campaign experts from across channels and teams under one umbrella. Trying to coordinate their work across separate silos would have created delays and disconnects as requests for input and approvals moved from one department to another. It would also have forced the individuals involved to juggle two sets of obligations.

Often AI project teams can simply be single squads, in which the whole

team carries out all the work by itself. But when the tasks are relatively broad in scope, requiring the work of more than a dozen people, a single team will be too unwieldy. In those situations it will make sense to divide the team into several squads, with one squad providing shared capabilities. The telecom company divided its new customer value team into four business squads—one focused on prepaid customers, one on postpaid customers, one on customer acquisition, and one on customer retention. It gave each a mission of either reducing churn or improving cross-sell by 20% by the end of the year. A fifth squad, data utility, with data engineers and developers, was created to support the other four by building technology and assets that could be reused by each one and by developing new AI-enabled analytics models.

STEP 3

Reimagine Business as Usual

As we noted earlier, getting the most from AI requires reinventing business models, roles and responsibilities, and operational processes using new ways of thinking and working. Typically, we find that companies are best served by applying first principles or design-thinking techniques and working backward from a key goal or challenge. For example, firms might envision what a five-star customer experience would look like and then explore in granular detail how they could achieve it.

At the airline the cargo team began by interviewing sales and reservation agents about how they allocated space on passenger planes and decided whether to accept or reject shipment requests. How did agents check on cargo space availability? What other information did they rely on, and how did they weigh the different pieces of information they collected? What concerns did they have when making decisions?

The team found that the legacy approach was plagued by inaccurate forecasts and guesswork by agents trying to estimate potential cancellations. (With cargo bookings, unlike passenger reservations, there's no penalty for canceling, so it's not unusual for a plane to look fully booked but leave with empty cargo space because of a no-show.) Cargo booking agents were also apprehensive about the impact on customer satisfaction if space was overbooked. To avoid conflicts, agents often waited until the day of the flight to book cargo space for their customers, resulting in suboptimal use of capacity and missed opportunities.

Having identified and understood the issues with the existing processes, the team then mapped out what an ideal process might look like, including the information that agents would need to determine whether to book, how much they could safely overbook and how far in advance, and how roles would be different. It then spent a few weeks developing a prototype of an AI-enabled dashboard that would provide the necessary information to agents, working in iterative sprints with them to incorporate input from the forecasting models, which were being developed in parallel. The team tested the dashboard with agents for 12 routes representative of the company's global network of 1,500. It compared how cargo utilization and profits differed on routes for agents who followed system recommendations and for a control group who used traditional processes. To build trust in the new system, executives eliminated any repercussions agents might normally face if a flight couldn't accommodate a reservation.

All agents now have access to intuitive dashboards that visually illustrate which flights are underutilizing space. They can view at a glance data on how cargo shipments for recent flights produced revenue. Integrated feedback loops enable the AI systems to continually learn from the agents as they decide whether to accept a cargo request, drawing on their expertise on shipment size and weight balance issues and their knowledge of changes in customers' supply chains, trade routes, and other factors. These new tools provide agents with information that gives them the confidence to sell cargo space well ahead of departure dates.

STEP 4

Adapt for Organizational and Technological Change

In most cases significant organizational changes, such as adopting interdisciplinary collaboration and agile mindsets, will be required to support the new AI-based processes and models. In fact, our research shows that the companies getting the highest returns on AI are more likely to enact effective change management practices, such as having leaders model desired behaviors, and

that such efforts work best when facilitated by CEOs and top executives.

Take the energy utility retailer again. It invested in reskilling employees so that they could effectively work together in the new context and take on new leadership responsibilities; realigned AI project team members' goals and incentives with their new responsibilities; and backfilled responsibilities in the departments the team members had to leave.

While companies will need to update their tech to support AI, they won't need to do major surgery on their IT infrastructure or data architecture before they begin. Rather, we advise companies to focus on technology that will enable and accelerate AI development and then triage additional investments according to teams' priorities. Cloud-based data platforms and the use of APIs, microservices, and other modern dev-ops practices, for example, can help companies develop new business capabilities two to three times faster.

The telecom provider established a cloud-based platform for raw data from existing transaction and customer service systems so that it could be used more easily by data engineers and data scientists than data from the old warehouse system could. The company also implemented a new analytics workbench, which helped the data scientists train and deploy new models faster, and tools that streamlined data collection, analysis, and model building for its AI-driven customer-value-management system. Those moves allowed it to begin using unstructured data, apply more complex approaches, and work more efficiently.

When prioritizing additional technology investments, teams should map out the capabilities, data, and resources (such as robotics, biometrics, and sensors and connectivity platforms) they will require and when, and then chip away at each piece as needed. In designing its customer-value-management system, the telecom provider's team realized it would need new technology that automated outbound direct messaging and gave salespeople real-time guidance about the next conversation to have with customers.

Teams should also consider the potential impact that AI initiatives will have on upstream and downstream processes and implement measures to address it. For example, at the airline the AI team developed a reporting tool for managers overseeing the loading and unloading of cargo so that they could effectively support the higher volumes produced by the new sales and reservation process.

A Domino Effect

Once AI development matures within an initial domain and organizations have gotten into a rhythm for reimagining parts of the business, they're ready to expand. The tech foundation they've built and the skills they've learned—for example, how to successfully break down silos, make decisions that used to take weeks in hours, and create more data-driven teams—will help accelerate their efforts in new domains.

At this point companies can pursue multiple domains in parallel. Again, the idea is to build off past work. This might lead companies to prioritize domains that have data and skills in common, such as supply chain and logistics. Or they might pursue the same domain in other business units. The energy utility retailer estimates that nearly 80% of the work done on improving customer value management in one product division (which led to record growth in just a few months, including a 12% increase in customer profit and a 20% increase in customer retention) can be reused in several other business units and accelerate their growth as well.

The companies profiled in this article are all still in the earlier stages of their full AI transformations, but they're on the threshold of a new era. They've gained a taste of what's possible, and their bold choices have yielded significant returns within the domains they've targeted and new capabilities that discrete use cases couldn't deliver. These companies have created a playbook of methodologies and protocols they can turn to again. As they move on to other domains, their pace will quicken, their AI capabilities will rapidly compound, and they'll find that the future they imagined is actually closer than it once appeared.

HBR Reprint R2103H

Tim Fountaine *is a senior partner in McKinsey's Sydney office and leads QuantumBlack, an advanced analytics firm owned by McKinsey, in Asia.* ***Brian McCarthy*** *is a partner for digital platforms and AI at private equity firm Chicago Pacific Founders and a former partner in McKinsey's Atlanta office leading analytics transformation.* ***Tamim Saleh*** *is a senior partner in McKinsey's London office and is a global analytics leader for McKinsey Digital.*

Raising the bar on *high-impact, custom* Executive Education

Delivered in various formats such as in-person, online, live-online, or in a hybrid learning experience.

SET YOUR ORGANIZATION APART - THINK CHICAGO BOOTH

Consistently ranked among the top programs globally, Executive Education at The University of Chicago Booth School of Business is where intellectual rigor, engaging teaching, and cutting-edge research meet. We work collaboratively with organizations to create custom design education programs and explore innovative virtual learning solutions to meet their unique learning objectives.

From top consumer packaged goods to global financial services, we've helped organizations bolstered their senior management's leadership, strategy, and financial decision-making skills. Working with both public and private sector clients, we've created general management programs for executives spanning the Americas, Asia, Europe, Africa, and the Middle East.

Our custom programs are delivered in an intense, exhilarating climate of discussion and debate. As a result, participants have transformational experiences and become empowered, impactful leaders. Partner with us and empower your organization with The Chicago Approach™.

PARTNER WITH CHICAGO BOOTH

Contact us to learn more about our customized approach.

+1 312.464.8732
exec.ed@chicagobooth.edu
ChicagoBooth.edu/Custom-Program

We offer various learning formats such as in-person, online, live-online, or blended.

CHICAGO BOOTH | Executive Education

CHICAGO LONDON HONG KONG

ORIGINALLY PUBLISHED ON HBR.ORG JULY 19, 2017

Inside Facebook's AI Workshop

At the social network behemoth, machine learning has become a platform for the platform.

→ by SCOTT BERINATO

WITHIN FACEBOOK'S CAVERNOUS Building 20, about halfway between the lobby (panoramic views of the Ravenswood Slough) and the kitchen (hot breakfast, smoothies, gourmet coffee), in a small conference room called Lollapalooza, Joaquin Candela is trying to explain artificial intelligence to a layperson.

Candela—bald, compact, thoughtful—runs Facebook's Applied Machine Learning (AML) group—the engine room of AI at Facebook, which, increasingly, makes it the engine room of Facebook in general. After some verbal searching, he finally says:

"Look, a machine learning algorithm really is a lookup table, right? Where the key is the input, like an image, and the value is the label for the input, like 'a horse.' I have a bunch of examples of something. Pictures of horses. I give the algorithm as many as I can. 'This is a horse. This is a horse. This isn't a horse. This is a horse.' And the algorithm keeps those in a table. Then, if a new example comes along—or if I tell it to watch for new examples—well, the algorithm just goes and looks at all those examples we fed it. Which rows in the table look similar? And how similar? It's trying to decide, 'Is this new thing a horse? I think so.' If it's right, the image gets put in the 'This is a horse' group, and if it's wrong, it gets put in the 'This isn't a horse' group. Next time, it has more data to look up.

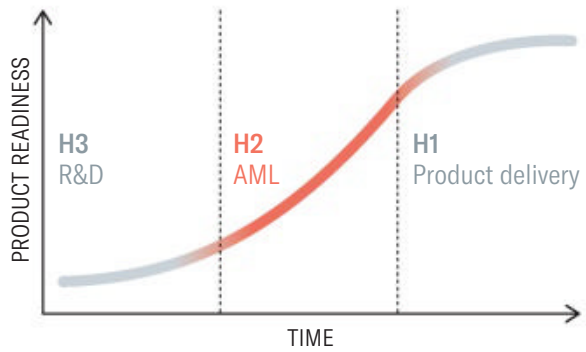
"One challenge is how do we decide how similar a new picture is to the ones stored in the table. One aspect of machine learning is to learn similarity functions. Another challenge is, What happens when your table grows really large? For every new image, you would need to make a zillion comparisons.... So another aspect of machine learning is to approximate a large stored table with a function instead of going through every image. The function knows how to roughly estimate what the corresponding value should be. That's the essence of ML—to approximate a gigantic table with a function. This is what learning is about."

There's more to it than that, obviously, but it's a good starting point when talking about AI, because it makes it sound real, almost boring. Mechanical. So much of the conversation around AI is awash in mystical descriptions of its power and in reverence for its near-magic capabilities. Candela doesn't like that and tries to use more-prosaic terms. It's powerful, yes, but not magical. It has limitations. During presentations, he's fond of showing a slide with a wizard and a factory, telling audiences

Illustration by ENISAURUS

Like
Like
Like

Where AI Fits In at Facebook



Source: Facebook

that Facebook thinks of AI like the latter, because "wizards don't scale."

And that's what Facebook has done with AI and machine learning: scaled it at a breakneck pace. A few years ago the company's machine learning group numbered just a few and needed days to run an experiment. Now, Candela says, several hundred employees run thousands of experiments a day. AI is woven so intricately into the platform that it would be impossible to separate the products—your feed, your chat, your kid's finsta—from the algorithms. Nearly everything users see and do is informed by AI and machine learning.

Understanding how and why Facebook has so fully embraced AI can help any organization that's ready to invest in an algorithmic future. It would be easy to assume that Facebook, with all its resources, would simply get the best talent and write the best algorithms—game over. But Candela took a different approach. Certainly the talent is strong, and the algorithms are good. Some of them are designed to "see" images or automatically filter them. Some understand conversations and can respond to them. Some translate between languages. Some try to predict what you'll like and buy.

But in some ways the algorithms are not his main focus. Instead, he's been busy creating an AI workshop in which anyone in the company can use AI to achieve a goal. Basically, Candela built an AI platform for the platform. Whether you're a deeply knowledgeable programmer or a complete newbie, you can take advantage of his wares.

Here's how he did it and what you can learn from it.

Soyuz

Candela, a veteran of Microsoft Research, arrived at Facebook in 2012 to work in the company's ads business. He and a handful of staffers inherited a ranking algorithm for better targeting users with ads.

Candela describes the machine learning code he inherited as "robust but not the latest." More than once he compares it to Soyuz, the 1960s Soviet spacecraft. Basic but reliable. Gets the job done even if it's not the newest, best thing. "It'll get you up there and down. But it's not the latest covnet [convolutional neural net] of the month."

You might assume, then, that the first thing Candela set out to do was to upgrade the algorithm. Get rid of Soyuz in favor of a space plane. It wasn't. "To get more value, I can do three things," he says. "I can improve the algorithm itself, make it more sophisticated. I can throw more and better data at the algorithm so that the existing code produces better results. And I can change the speed of experimentation to get more results faster.

"We focused on data and speed, not on a better algorithm."

Candela describes this decision as "dramatic" and "hard." Computer scientists, especially academic-minded ones, are rewarded for inventing new algorithms or improving existing ones. A better statistical model is the goal. Getting cited in a journal is validation. Wowing your peers gives you cred.

It requires a shift in thinking to get those engineers to focus on business impact before optimal statistical model. He thinks many companies are making the mistake of structuring their efforts around building the best algorithms, or hiring developers who claim to have the best algorithms, because that's how many AI developers think.

But for a company, a good algorithm that improves the business is more valuable than vanguard statistical models. In truth, Candela says, real algorithmic breakthroughs are few and far between—two or three a year at best. If his team focused its energies there, it would take lots of effort to make marginal gains.

He hammers these points home constantly: Figure out the impact on the business first. Know what you're solving for. Know what business challenge you need to address. "You might look for the shiniest algorithm or the people who are telling you they have the most advanced algorithm. And you really should be looking for people who are most obsessed with getting any algorithm to do

It requires a shift in thinking to get computer scientists to focus on business impact before optimal statistical models.

a job. That's kind of a profound thing that I think is lost in a lot of the conversation. I had a conversation with our resident machine learning geek at our office, and we were just talking about different people doing AI. He said, 'Nobody really thinks their algorithms are very good or whatever.' It makes me think, maybe that's fine.

"I'm not saying don't work on the algorithm at all. I'm saying that focusing on giving it more data and better data, and then experimenting faster, makes a lot more sense."

So rather than defining success as building the best natural language processing algorithm, he defines it as deploying one that will help users find a restaurant when they ask their friends, "Where can I get a good bite around here?" Instead of being thrilled that some computer vision algorithm is nearing pixel-perfect object recognition, he gets excited if that AI is good enough to notice that you post a lot of pictures of the beach and can help you buy a swimsuit.

The strategy worked when he started at Facebook. Ad revenues rose. Candela's profile rose. It was suggested that AML become a centralized function for all of Facebook. Candela said no. Twice. "I was concerned about the 'If you build it, they will come' phenomenon." Just creating bits of artificial intelligence in the hope that people would see the value and adopt it wouldn't work.

But he did pick his spots. He collaborated with the feeds team while saying no to many other groups. Then he worked with the Messenger team. His team grew and took on more projects with other teams.

By 2015 Candela could see that his group would need to centralize, so he turned his attention to how he'd build such an operation. He was still worried about the "build it and they will come" phenomenon, so he focused less on how his team would be structured and more on how the group would connect to the rest of Facebook. "You build a factory that makes amazing widgets, and you forget to design the loading docks into your factory?" He laughs. "Well, enjoy your widgets."

Only then, about three years in, did Candela think about upgrading some of his algorithms. (Incidentally, even today, the emergency escape spacecraft attached to the International Space Station is a Soyuz.)

H2

Candela goes to a whiteboard to describe how he built his AI factory inside Facebook. The key, he says, was figuring out where on the product development path AI fits. He draws something like the graph in the exhibit "Where AI Fits In at Facebook."

H3—Horizon 3 or three years out from product—is the realm of R&D and science. Often, data scientists who work on AI think of themselves as here, improving algorithms and looking for new ways to get machines to learn. Candela didn't put his team here for the reasons already mentioned. It's too far from impact on the business. H1, approaching product delivery, is where the product teams live—the feeds team, the Instagram team, the ads team. AI doesn't go here either, because it would be difficult to retrofit products this deeply developed.

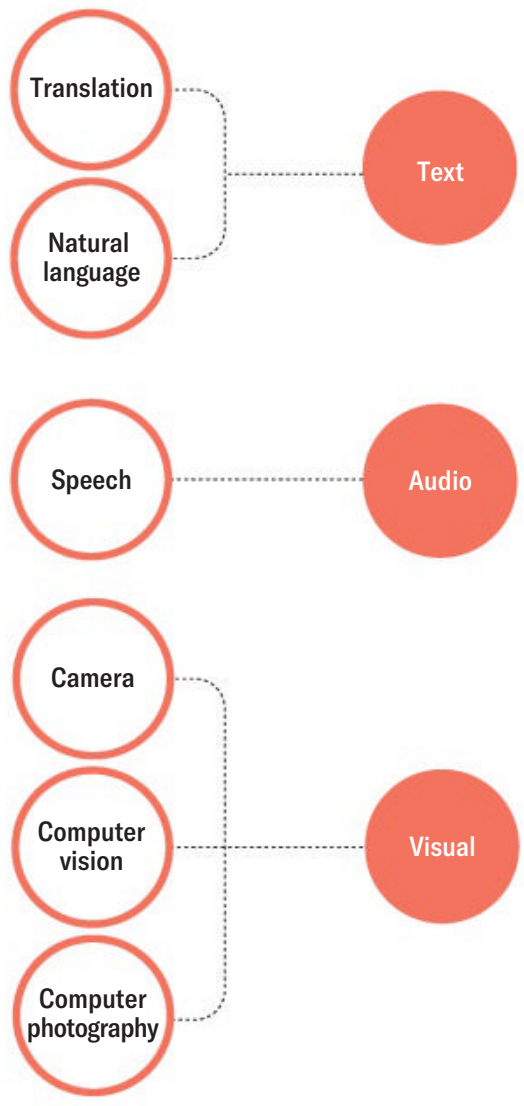
Idea in Brief

THE MYTHICAL MAGIC BULLET
Much of the conversation around AI is filled with reverential descriptions of its power and near-magical capabilities. But AI is most useful to organizations when it's harnessed for a large number of much more prosaic tasks.

AI AT FACEBOOK
Understanding how and why Facebook has embraced AI can help any organization ready to invest in an algorithmic future. Joaquin Candela runs Facebook's Applied Machine Learning (AML) group. Rather than focusing on the shiniest, most cutting-edge algorithms, he and his team have built a platform that allows their colleagues to create workaday algorithms that actually serve the needs of the business.

FOCUS ON THE PRACTICAL
Candela and Facebook's AML group take as a measure of their success whether their AI is genuinely helping users, rather than whether it has the best natural-language processing. Can it actually help users find a restaurant where they can meet their friends? Can it recommend the right swimsuit if it notices them posting pictures of the beach?

Applied Machine Learning



Source: Facebook

It would be like building a car and then deciding that it should be self-driving after you started to put it together.

That leaves H2, between the science and the product, as the place AML lives at Facebook. AML is a conduit for transferring the science into the product. It does not do research for research's sake, and it does not build and ship products. As the upward slope in the product's readiness shows, it's a dynamic space. Pointing to H2, Candela says, "This needs to feel uncomfortable all the time. The people you need to hire need to be okay with that, and they need to be incredibly selfless. Because if your work is successful, you spin it out. And you need to fail quite a bit. I'm comfortable with a 50% failure rate."

If the team is failing less, Candela suspects its members are too risk averse, or they're taking on challenges that are sliding them closer to H1's product focus. "Maybe we do something like that and it works, but it's still a failure, because the product teams should be taking that on, not us. If you own a piece of technology that the ads team should operate themselves to generate value, give it to them, and then increase your level of ambition in the machine learning space before something becomes product."

So Candela's team is neither earning the glory of inventing new statistical models nor putting products out into the world. It's a factory of specialists who translate others' science for others' products and fail half the time.

Push/Pull

All that being said, the lines between the three realms—H3, H2, and H1—still aren't crisp. In some cases Candela's team does look at the science of machine learning, to solve specific problems. And sometimes it does help build the products.

That was especially true as AML got off the ground, because many people in the business hadn't yet been exposed to AI and what it could do for them. In one case AML built a translation algorithm. The team dipped into the research space to look at how existing translation algorithms worked and could be improved, because bad translations, which either don't make sense or create a misleading interpretation, are in some ways worse than no translation.

"Early on it was more push, more tenacity on our part," Candela says. "But it was gentle tenacity. We weren't going to throw something over the fence and tell the product team, 'This is great, use it.'" That meant that his team helped write some product code. Doing a little bit of the science and a little bit of the product in addition to its core function was meant to inspire the product team members to see what AML could do for them.

What the two teams built—a product that allowed community pages to instantly translate into several languages—worked. Other projects were similarly pushed out, and now the international team and other product groups at Facebook are pulling from AML, asking to use code in their products themselves.

"Look, it's nowhere near where I want it to be," Candela says. "I'd like to have all the product leaders in the company get together quarterly for AI

Figure out the impact on the business first. Know what you're solving for. Know what business challenge you need to address.

reviews. That will certainly happen. But the conversation in the past two years has completely changed. Now if I walk from one end of this building to the other and I bump into, I don't know, the video team or the Messenger team, they'll stop me and say, 'Hey, we're excited to try this. We think we can build a product on this.' That didn't happen before."

AML's success, though, has created a new challenge for Candela. Now that everyone wants a piece of AML, the factory has to scale.

Layer Cake

Candela couldn't scale just by saying yes to every project and adding bodies to get the work done. So he organized in other ways. First he subdivided his team according to the type of AI its members would focus on. (See the exhibit "Applied Machine Learning.")

This created common denominators so that one team—say, computer vision—could work on any machine learning application involving parsing images and reuse its work whenever possible.

Next came a large-scale engineering effort to build Facebook's own AI backbone, called FBLearner Flow. Here algorithms are deployed once and made reusable for anyone who may need them. The time-consuming parts of setting up and running experiments are automated, and past results are stored and made available and easily searchable. And the system runs through a serious hardware array, so many experiments can be run simultaneously. (The system allows for more than 6 million predictions a second). All of this is to increase the velocity of running experiments on the data and scale.

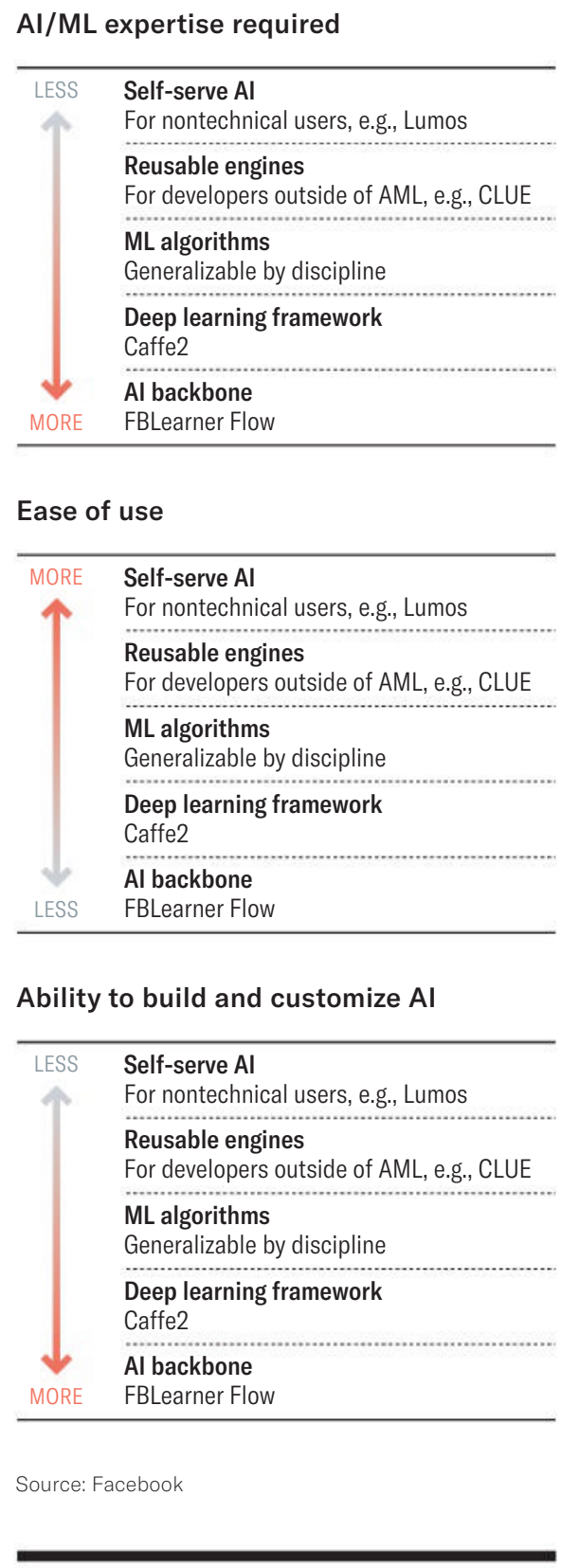
The system was also designed to accommodate many kinds of possible users. Candela believes that for AI to work, and to scale even further, he must help people outside AML do the work themselves. He created what he calls a layer cake of artificial intelligence (see the exhibit at right).

The bottom layers focus on AML's work: refining the core engine (with a strong focus on optimizing performance, especially for mobile) and working with machine learning algorithms. The upper layers focus on tools that make it possible for those outside AML to exploit the algorithms with less AML involvement. "It's all about what you expose to the user," Candela says. In some cases he's built systems that developers outside AML can take advantage of to build and run their own models.

Rex

A good example of Candela's team structure and the push/pull dynamic comes from some AI built to surface content on the basis of what you type. The natural-language machine learning team created an engine to understand conversational typing.

This bit of intelligence first found its way into the Messenger chat client. AML developed the models while the product team developed use cases and "intents"—lingo for the types of tasks you want the engine to learn. For example, training natural language AI to recognize and reliably respond to a



Source: Facebook

"I tell people, 'You don't need machine learning. You need a data science team to think through a problem and apply the human litmus test.' Exhaust the human brain first."

phrase like "I'm looking for the best..." is an intent.

The first few such intents were deployed to Messenger through a product called M Suggestions.

If you sent a chat to a friend that said "I'll meet you there in 30 minutes," M Suggestions might prompt you with an offer to hire a car.

As the tools for building intent models developed and the product team became more conversant with them, AML's role diminished. Now the Messenger team has improved M Suggestions by building dozens more intents on its own.

Still, this bit of natural language AI wasn't built just for chat. It's reusable. It was codified as CLUE, for "conversational learning understanding engine." It found its way into more Facebook applications. It's being adapted for status updates and feeds. Social recommendations—or social rex, as everyone calls them—are increasingly driven by AI. If you typed "I'm traveling to Omaha and I really want to find a good steak downtown," AI might respond as if it were one of your friends, with a comment on your post, rex such as a list of steakhouses, and a tagged map of where they are relative to downtown. If your friend replied to you and said, "It also has some great vegetarian restaurants," the algorithm might again reply with pertinent data.

Social rex intents are not yet being developed without AML, but the goal is to have them move out of Candela's group, just as M Suggestions did.

In general, the idea is to make product teams AI-capable themselves. "We'll teach you to fish," Candela says, "and you go fish, and we'll drag up the next thing. We'll build a fishing boat. And once you're using the fishing boat, I'm going to build a cannery, right?"

At the moment, about 70% of the AI work on the backbone is done by people outside Candela's team. That's possible in part because of the interface with AI. In some cases, as with a tool called Lumos, machine learning can be used by nondevelopers.

Horseback Riding and Cereal Boxes

Lumos is computer vision AI, a tool that can comb through photos on Facebook or Instagram or other platforms and learn what they contain. You can train it to see anything. It has helped automate the discovery and banning of pornographic or violent content, IP appropriation (improper use of brands and logos), and other unwelcome content. It can also help identify things you like and do (to drive personalized advertising and recommendations), on the basis of photos in your feeds.

I watch a demo in which engineers select "horseback riding" as our intent, the thing we'll be looking for. The interface is simple: A few clicks, a couple of forms to fill out—What are you looking for? How much data do you want to look at?—and the algorithm gets to work finding pictures of horseback riding. Thumbnails start to fill the page.

The algorithm has searched for horseback riding before, so it's already quite good at finding it. My guess is that north of 80% of the images that pop up are indeed of horseback riding, and they show remarkable variety. Here's one with someone posing at a standstill. Here's one with the horse rearing. Here's an equestrian jumping. The algorithm finds shapes and boundaries between shapes and builds on previous knowledge of what those interactions mean. It knows things about what combination of pixels is most likely a person, for example, and what's a horse. It knows when it "sees" a person and a horse together with the person situated close above the horse. And it decides that this looks like horseback riding.

We also find pictures that aren't horseback riding—one is a person standing next to a horse; another is a person on a mule—and check those off as not matches. They're framed in red, in case there's any doubt. The algorithm internalizes that information—adds it to the lookup table—for use next time. A simple chart at the top of the page shows the algorithm's accuracy and confidence over time. It's always an S curve, slow to learn at first, then rapidly improving, then tapering off on how much more accurate it can get. It's very good at seeing horseback riding.

Other potentially valuable pictures are harder for AI to parse. "Receipts" is tricky to suss out because it can look to a computer just like type on a page; but there would be some interesting apps for AI that could recognize and "read" receipts. The engineers show how bowling alleys and escalators often confuse the algorithm because they have similar shapes and visual properties.

I ask, "What about something like 'food'?" This brings us to an important point about machine learning: It's only as good as its training.

We call up food as a topic to train. Indeed, we see lots of pictures of fruits and vegetables, a few of plates at restaurants. All food. We also see a cereal box. Is that food?

Well, yes. Or no. It's a box. But there's food in it. When we buy it, we're buying food, not the box. If I asked if there was any food in the cupboard, you wouldn't say, "No, just a cereal box." (Or, more pertinent to Facebook, if I posted a picture of a cereal box, should it think I'm posting about food or about a box?) As a picture, as a piece of data, it's a box.

Should we mark this as a match or a miss? Here's part of the art of machine learning. When training algorithms, one needs to use clearly definable categories. Food is probably too general in some ways, and the algorithm will either improperly hit or miss on images because it's hard to know what we mean when we say, "Show me pictures of food." "Vegetable" is a better idea to train on. And when training, everyone must define terms in the same way. Imagine two people training the algorithm when one always marks cereal boxes as food, and the other marks them as not food. Now imagine that happening at scale, on terabytes of visual data.

The same applies to natural language processing. Humans are very good at interpreting text in context to find sophisticated meaning. For example, I may type, "Gee, I love that movie about the superheroes. It's so, so original! I hope they make a hundred more of them." My friends, who know me and know some of the mechanics of sarcasm, may readily understand that my meaning is the opposite of what I'm typing. Artificial intelligence is still learning how to decide the meaning of something like that. To figure out if I'm being sarcastic, it has to go much further than just learning how to parse grammar and vocabulary. It has to see what else I've said and posted, and try to find other clues that will tell it whether I really loved the movie and I want 100 more or I actually detested it—because getting that wrong is not good for a platform that wants to create affinities with me. If I was being sarcastic and my feed starts filling up with superhero movie ads, I'm probably not enjoying the experience.

Not Magic

It's details like these—showing where AI is still limited, and how humans have such a core role in training it, and how solving problems and creating value are more important than finding great models—that Candela is thinking about near the end of the day, when he's talking about the mythic status AI has gained. He's railing against what he perceives as laziness in those who find the idea of AI-as-magic-bullet appealing and don't apply critical thinking to it.

"What frustrates me," he says, "is that everybody knows what a statistician is and what a data analyst can do. If I want to know 'Hey, what age segment behaves in what way?' I get the data analyst.

"So when people skip that, and they come to us and say, 'Hey, give me a machine learning algorithm that will do what we do,' I'm like, 'What is it that I look like? What problem are you trying to solve? What's your goal? What are the trade-offs?'" Sometimes they're surprised that there are trade-offs. "If that person doesn't have answers to those questions, I'm thinking, 'What the hell are you thinking AI is?'"

They are thinking it's magic.

"But it's not. That's the part where I tell people, 'You don't need machine learning. You need to build a data science team that helps you think through a problem and apply the human litmus test. Sit with them. Look at your data. If you can't tell what's going on, if you don't have any intuition, if you can't build a very simple, rule-based system—like, *Hey, if a person is younger than 20 and living in this geography, then do this thing*—if you can't do that, then I'm extremely nervous even talking about throwing AI at your problem.'

"I'm delighted when other executives come to me and start not from wanting to understand the technology but from a problem they have that they've thought very, very deeply about. And sometimes—often, in fact—a simple, good old rule-based system, if you have the right data, will get you 80% of the way to solving the problem.

"And guess what? It's going to have the benefit that everybody understands it. Exhaust the human brain first."

HBR Reprint H03QY6
Big Idea Package Reprint BG1704

Scott Berinato *is a senior editor at* Harvard Business Review *and the author of* Good Charts: The HBR Guide to Making Smarter, More Persuasive Data Visualizations *(2016) and* Good Charts Workbook: Tips, Tools, and Exercises for Making Better Data Visualizations *(2019).*

HI!

ORIGINALLY PUBLISHED JULY–AUGUST 2020

A Better Way to Onboard AI

Understand it as a tool to assist rather than replace people.

→ by BORIS BABIC, DANIEL L. CHEN, THEODOROS EVGENIOU, and ANNE-LAURE FAYARD

IN A 2018 WORKFORCE INSTITUTE survey of 3,000 managers across eight industrialized nations, the majority of respondents described artificial intelligence as a valuable productivity tool.

It's easy to see why: AI brings tangible benefits in processing speed, accuracy, and consistency (machines don't make mistakes because they're tired), which is why many professionals now rely on it. Some medical specialists, for example, use AI tools to help make diagnoses and decisions about treatment.

But respondents to that survey also expressed fears that AI would take their jobs. They are not alone. The *Guardian* recently reported that more than 6 million workers in the UK fear being replaced by machines. These fears are echoed by academics and executives we meet at conferences and seminars. AI's advantages can be cast in a much darker light: Why would humans be needed when machines can do a better job?

The prevalence of such fears suggests that organizations looking to reap the benefits of AI need to be careful when introducing it to the people expected to work with it. Andrew Wilson, until January 2020 Accenture's CIO, says, "The greater the degree of organizational focus on people helping AI, and AI helping people, the greater the value achieved." Accenture has found that

Illustration by **ENISAURUS**

when companies make it clear that they are using AI to help people rather than to replace them, they significantly outperform companies that don't set that objective (or are unclear about their AI goals) along most dimensions of managerial productivity—notably speed, scalability, and effectiveness of decision-making.

In other words, just as when new talent joins a team, AI must be set up to succeed rather than to fail. A smart employer trains new hires by giving them simple tasks that build hands-on experience in a noncritical context and assigns them mentors to offer help and advice. This allows the newcomers to learn while others focus on higher-value tasks. As they gain experience and demonstrate that they can do the job, their mentors increasingly rely on them as sounding boards and entrust them with more-substantive decisions. Over time an apprentice becomes a partner, contributing skills and insight.

We believe this approach can work for artificial intelligence as well. In the following pages we draw on our own and others' research and consulting on AI and information systems implementation, along with organizational studies of innovation and work practices, to present a four-phase approach to implementing AI. It allows enterprises to cultivate people's trust—a key condition for adoption—and to work toward a distributed human-AI cognitive system in which people and AI *both* continually improve. Many organizations have experimented with phase 1, and some have progressed to phases 2 and 3. For now, phase 4 may be mostly a "future-casting" exercise of which we see some early signs, but it is feasible from a technological perspective and would provide more value to companies as they engage with artificial intelligence.

PHASE 1

The Assistant

This first phase of onboarding artificial intelligence is rather like the process of training an assistant. You teach the new employee a few fundamental rules and hand over some basic but time-consuming tasks you normally do (such as filing online forms or summarizing documents), which frees you to focus on more-important aspects of the job. The trainee learns by watching you, performing the tasks, and asking questions.

One common task for AI assistants is sorting data. An example is the recommendation systems companies have used since the mid-1990s to help customers filter thousands of products and find the ones most relevant to them—Amazon and Netflix being among the leaders in this technology.

More and more business decisions now require this type of data sorting. When, for example, portfolio managers are choosing stocks in which to invest, the information available is far more than a human can feasibly process, and new information comes out all the time, adding to the historical record. Software can make the task more manageable by immediately filtering stocks to meet predefined investment criteria. Natural-language processing, meanwhile, can identify the news most relevant to a company and even assess the general sentiment about an upcoming corporate event as reflected in analysts' reports. Marble Bar Asset Management (MBAM), a London-based investment firm founded in 2002, is an early convert to using such technologies in the workplace. It has developed a state-of-the-art platform, called RAID (Research Analysis & Information Database), to help portfolio managers filter through high volumes of information about corporate events, news developments, and stock movements.

Another way AI can lend assistance is to model what a human might do. As anyone who uses Google will have noticed, prompts appear as a search phrase is typed in. Predictive text on a smartphone offers a similar way to speed up the process of typing. This kind of user modeling, related to what is sometimes called *judgmental bootstrapping,* was developed more than 30 years ago; it can easily be applied to decision-making. AI would use it to identify the choice an employee is most likely to make, given that employee's past choices, and would suggest that choice as a starting point when the employee is faced with multiple decisions—speeding up, rather than actually doing, the job.

Let's look at this in a specific context. When airline employees are deciding how much food and drink to put on a given flight, they fill out catering orders, which involve a certain amount of calculation together with assumptions based on their experience of previous flights. Making the wrong choices incurs costs: Underordering risks upsetting customers who may avoid future travel on the airline. Overordering means the excess food will go to waste

AI is especially helpful during high-volume decision-making, when human employees may be tired or distracted.

and the plane will have increased its fuel consumption unnecessarily.

An algorithm can be very helpful in this context. AI can predict what the airline's catering manager would order by analyzing his or her past choices or using rules set by the manager. This "autocomplete" of "recommended orders" can be customized for every flight using all relevant historical data, including food and drink consumption on the route in question and even past purchasing behavior by passengers on the manifest for that flight. But as with predictive typing, human users can freely overwrite as needed; they are always in the driver's seat. AI simply assists them by imitating or anticipating their decision style.

It should not be a stretch for managers to work with AI in this way. We already do so in our personal lives, when we allow the autocomplete function to prefill forms for us online. In the workplace a manager can, for example, define specific rules for an AI assistant to follow when completing forms. In fact, many software tools currently used in the workplace (such as credit-rating programs) are already just that: collections of human-defined decision rules. The AI assistant can refine the rules by codifying the circumstances under which the manager actually follows them. This learning needn't involve any change in the manager's behavior, let alone any effort to "teach" the assistant.

PHASE 2

The Monitor

The next step is to set up the AI system to provide real-time feedback. Thanks to machine-learning programs, AI can be trained to accurately forecast what a user's decision would be in a given situation (absent lapses in rationality owing to, for example, overconfidence or fatigue). If a user is about to make a choice that is inconsistent with his or her choice history, the system can flag the discrepancy. This is especially helpful during high-volume decision-making, when human employees may be tired or distracted.

Research in psychology, behavioral economics, and cognitive science shows that humans have limited and imperfect reasoning capabilities, especially when it comes to statistical and probabilistic problems, which are ubiquitous in business. Several studies (of which one of us, Chen, is a coauthor) concerning legal decisions found that judges grant political asylum more frequently before lunch than after, that they give lighter prison sentences if their NFL team won the previous day than if it lost, and that they will go easier on a defendant on the latter's birthday. Clearly justice might be better served if human decision makers were assisted by software that told them when a decision they were planning to make was inconsistent with their prior decisions or with the decision that an analysis of purely legal variables would predict.

AI can deliver that kind of input. Another study (also with Chen as a coauthor) showed that AI programs processing a model made up of basic legal variables (constructed by the study's authors) can predict asylum decisions with roughly 80% accuracy on the date a case opens. The authors have added learning functionality to the program, which enables it to simulate the decision-making of an individual judge by drawing on that judge's past decisions.

The approach translates well to other contexts. For example, when portfolio managers (PMs) at Marble Bar Asset Management consider buy or sell decisions that may raise the overall portfolio risk—for example, by increasing exposure to a particular sector or geography—the system alerts them through a pop-up during a computerized transaction process so that they can adjust appropriately. A PM may ignore

Idea in Brief

THE PROBLEM
Many companies struggle to apply AI and fail to achieve the productivity improvements they seek.

WHY IT HAPPENS
Executives often don't make clear that they are using AI to help people increase productivity rather than to replace them.

HOW TO FIX IT
Treat AI adoption as an onboarding process that consists of four phases: AI as an assistant, as a monitor, as a coach, and as a teammate.

Of course AI is not always "right." That's why using it should be like a dialogue, to improve its usefulness and preserve the autonomy of the human decision maker.

such feedback as long as company risk limits are observed. But in any case the feedback helps the PM reflect on his or her decisions.

Of course AI is not always "right." Often its suggestions don't take into account some reliable private information to which the human decision maker has access, so the AI might steer an employee off course rather than simply correct for possible behavioral biases. That's why using it should be like a dialogue, in which the algorithm provides nudges according to the data it has while the human teaches the AI by explaining why he or she overrode a particular nudge. This improves the AI's usefulness and preserves the autonomy of the human decision maker.

Unfortunately, many AI systems are set up to usurp that autonomy. Once an algorithm has flagged a bank transaction as possibly fraudulent, for example, employees are often unable to approve the transaction without clearing it with a supervisor or even an outside auditor. Sometimes undoing a machine's choice is next to impossible—a persistent source of frustration for both customers and customer service professionals. In many cases the rationale for an AI choice is opaque, and employees are in no position to question that choice even when mistakes have been made.

Privacy is another big issue when machines collect data on the decisions people make. In addition to giving humans control in their exchanges with AI, we need to guarantee that any data it collects on them is kept confidential. A wall ought to separate the engineering team from management; otherwise employees may worry that if they freely interact with the system and make mistakes, they might later suffer for them.

Also, companies should set rules about designing and interacting with AI to ensure organizational consistency in norms and practices. These rules might specify the level of predictive accuracy required to show a nudge or to offer a reason for one; criteria for the necessity of a nudge; and the conditions under which an employee should either follow the AI's instruction or refer it to a superior rather than accept or reject it.

To help employees retain their sense of control in phase 2, we advise managers and systems designers to involve them in design: Engage them as experts to define the data that will be used and to determine ground truth; familiarize them with models during development; and provide training and interaction as those models are deployed. In the process, employees will see how the models are built, how the data is managed, and why the machines make the recommendations they do.

PHASE 3

The Coach

In a recent PwC survey nearly 60% of respondents said that they would like to get performance feedback on a daily or a weekly basis. It's not hard to see why. As Peter Drucker asserted in his famous 2005 *Harvard Business Review* article "Managing Oneself," people generally don't know what they are good at. And when they think they do know, they are usually wrong.

The trouble is that the only way to discover strengths and opportunities for improvement is through a careful analysis of key decisions and actions. That requires documenting expectations about outcomes and then, nine months to a year later, comparing those expectations with what actually happened. Thus the feedback employees get usually comes from hierarchical superiors during a review—not at a time or in a format of the recipient's choosing. That is unfortunate, because, as Tessa West of New York University found in a recent neuroscience study, the more people feel that their autonomy is protected and that they are in control of the conversation—able to choose, for example, when feedback is given—the better they respond to it.

AI could address this problem. The capabilities we've already mentioned could easily generate feedback for employees, enabling them to look at their own performance and reflect on variations and errors. A monthly summary analyzing data drawn from their past behavior might help them better understand their decision patterns and practices. A few companies, notably in the financial sector, are taking this approach. Portfolio managers at MBAM, for example, receive feedback from a data analytics system that captures investment decisions at the individual level.

The data can reveal interesting and varying biases among PMs. Some may be more loss-averse than others, holding on to underperforming investments longer than they should. Others may be overconfident, possibly taking on too large a position in a given investment. The analysis identifies these behaviors and—like a coach—provides personalized feedback that highlights behavioral

When AI Loses Its Way

In 2016 the investigative newsroom *ProPublica* published an exposé of a risk-prediction AI program known as COMPAS, which judges in southern Florida use to determine a defendant's likelihood of re-offending within a specified time period.

The algorithm underlying COMPAS is held as a trade secret by its manufacturer, Northpointe (now Equivant), which means that we don't know how COMPAS generates its predictions, nor do we have access to the data the algorithm is trained on—so we cannot even inquire into its rationale. When it was reported that the algorithm produces disparate outcomes across race, COMPAS immediately became a leading example of why people cannot trust AI.

If businesses want employees to adopt, use, and ultimately trust AI systems, it will be important to open up the black box—to the extent legally possible—to those who are expected to engage with the technology. As Richard Socher, the chief scientist at Salesforce, puts it, "If businesses use AI to make predictions, they owe humans an explanation as to how the decisions are made."

changes over time, suggesting how to improve decisions. But it is up to the PMs to decide how to incorporate the feedback. MBAM's leadership believes this "trading enhancement" is becoming a core differentiator that both helps develop portfolio managers and makes the organization more attractive.

What's more, just as a good mentor learns from the insights of the people who are being mentored, a machine-learning "coachbot" learns from the decisions of an empowered human employee. In the relationship we've described, a human can disagree with the coachbot—and that creates new data that will change the AI's implicit model. For example, if a portfolio manager decides not to trade a highlighted stock because of recent company events, he or she can provide an explanation to the system. With feedback, the system continually captures data that can be analyzed to provide insights.

If employees can relate to and control exchanges with artificial intelligence, they are more likely to see it as a safe channel for feedback that aims to help rather than to assess performance. Choosing the right interface is useful to this end. At MBAM, for example, trading enhancement tools—visuals, for instance—are personalized to reflect a PM's preferences.

As in phase 2, involving employees in designing the system is essential. When AI is a coach, people will be even more fearful of disempowerment. It can easily seem like a competitor as well as a partner—and who wants to feel less intelligent than a machine? Concerns about autonomy and privacy may be even stronger. Working with a coach requires honesty, and people may hesitate to be open with one that might share unflattering data with the folks in HR.

Deploying AI in the ways described in the first three phases does of course have some downsides. Over the long term new technologies create more jobs than they destroy, but meanwhile labor markets may be painfully disrupted. What's more, as Matt Beane argues in "Learning to Work with Intelligent Machines" (HBR, September–October 2019), companies that deploy AI can leave employees with fewer opportunities for hands-on learning and mentorship.

There is some risk, therefore, not only of losing entry-level jobs (because digital assistants can effectively replace human ones) but also of compromising the ability of future decision makers to think for themselves. That's not inevitable, however. As Beane suggests, companies could use their artificial intelligence to create different and better learning opportunities for their employees while improving the system by making it more transparent and giving employees more control. Because future entrants to the workforce will have grown up in a human-plus-machine workplace, they will almost certainly be faster than their pre-AI colleagues at spotting opportunities to innovate and introduce activities that add value and create jobs—which brings us to the final phase.

PHASE 4
The Teammate

Edwin Hutchins, a cognitive anthropologist, developed what is known as the theory of distributed cognition. It is based on his study of ship navigation, which, he showed, involved a combination of sailors, charts, rulers, compasses, and a plotting tool. The theory broadly relates to the concept of extended mind, which posits that cognitive processing, and associated mental acts such as belief and intention, are not necessarily limited to the brain, or even the body. External tools and instruments can, under the right conditions, play a role in cognitive processing and create what is known as a *coupled system*.

In line with this thinking, in the final phase of the AI implementation journey (which to our knowledge no organization has yet adopted) companies would develop a coupled network of humans and machines in which both contribute expertise. We believe that as AI improves through its interactions with individual users, analyzing and even modeling expert users by drawing on data about their past decisions and behaviors, a community of experts (humans and machines) will naturally emerge in organizations that have fully integrated AI coachbots. For example, a purchasing manager who—with one click at the moment of decision—could see what price someone else would give could benefit from a customized collective of experts.

Although the technology to create this kind of collective intelligence now exists, this phase is fraught with challenges. For example, any such integration of AI must avoid building in old or new biases and must respect human privacy concerns so that people can trust the AI as much as they would a human partner. That in itself is a pretty big challenge, given the volume of research demonstrating how hard it is to build trust among humans.

The best approaches to building trust in the workplace rely on the relationship between trust and understanding—a subject of study by David Danks and colleagues at Carnegie Mellon. According to this model, I trust someone because I understand that person's values, desires, and intentions, and they demonstrate that he or she has my best interests at heart. Although understanding has historically been a basis for building trust in human relationships, it is potentially well suited to cultivating human–AI partnerships as well, because employees' fear of artificial intelligence is usually grounded in a lack of understanding of how AI works. (See the sidebar "When AI Loses Its Way.")

In building understanding, a particular challenge is defining what "explanation" means—let alone "good explanation." This challenge is the focus of a lot of research. For example, one of us (Evgeniou) is working to open up machine-learning "black boxes" by means of so-called counterfactual explanations. A counterfactual explanation illuminates a particular decision of an AI system (for example, to approve credit for a given transaction) by identifying a short list of transaction characteristics that drove the decision one way or another. Had any of the characteristics been different (or counter to the fact), the system would have made a different decision (credit would have been denied).

Evgeniou is also exploring what people perceive as good explanations for AI decisions. For example, do they see an explanation as better when it's presented in terms of a logical combination of features ("The transaction was approved because it had X,Y,Z characteristics") or when it's presented relative to other decisions ("The transaction was approved because it looks like other approved transactions, and here they are for you to see")? As research into what makes AI explainable continues, AI systems should become more transparent, thus facilitating trust.

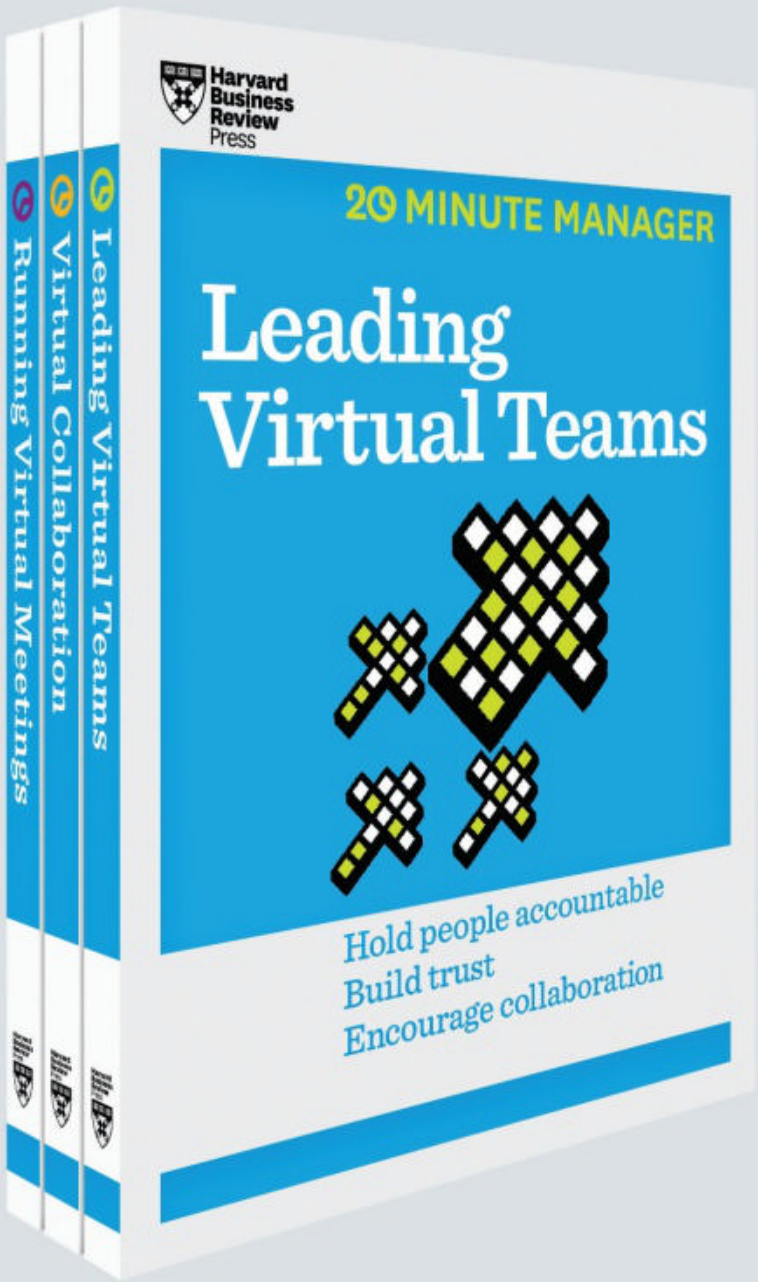
ADOPTING NEW TECHNOLOGIES has always been a major challenge—and the more impact a technology has, the bigger the challenge is. Because of its potential impact, artificial intelligence may be perceived as particularly difficult to implement. Yet if done mindfully, adoption can be fairly smooth. That is precisely why companies must ensure that AI's design and development are responsible—especially with regard to transparency, decision autonomy, and privacy—and that it engages the people who will be working with it. Otherwise they will quite reasonably fear being constrained—or even replaced—by machines that are making all sorts of decisions in ways they don't understand.

Getting past these fears to create a trusting relationship with AI is key. In all four phases described in these pages, humans determine the ground rules. With a responsible design, AI may become a true partner in the workplace—rapidly processing large volumes of varied data in a consistent manner to enhance the intuition and creativity of humans, who in turn teach the machine.

HBR Reprint R2004C

Boris Babic *is an assistant professor of philosophy and of statistics at The University of Toronto.* ***Daniel L. Chen*** *is a professor at the Institute for Advanced Study at the Toulouse School of Economics and lead investigator at the World Bank's Data and Evidence for Justice Reform program.* ***Theodoros Evgeniou*** *is a professor of decision sciences and technology management at INSEAD and an adviser to Marble Bar Asset Management (an investment firm named in this article).* ***Anne-Laure Fayard*** *is the ERA Chaired Professor in Social Innovation, NOVA School of Business and Economics.*

Harvard Business Review



Be Productive, No Matter Where You Are

Effectively collaborate and manage remotely while completing your own work more efficiently.

Get practical guidance as you work with colleagues in different locations, set shared expectations for your interactions, communicate successfully through a variety of media, and run productive virtual meetings. This collection includes:

- Three ebooks on virtual collaboration, running virtual meetings, and leading virtual teams
- A manual offering step-by-step guidance
- Customizable worksheets and templates
- Checklists, sample scripts, and other tips to improve virtual collaboration
- Five bonus HBR articles that dive deeper into best practices
- An HBR video about building trust on a virtual team

DOC PDF

INCLUDES CUSTOMIZABLE TOOLS AND TEMPLATES IN MULTIPLE FORMATS

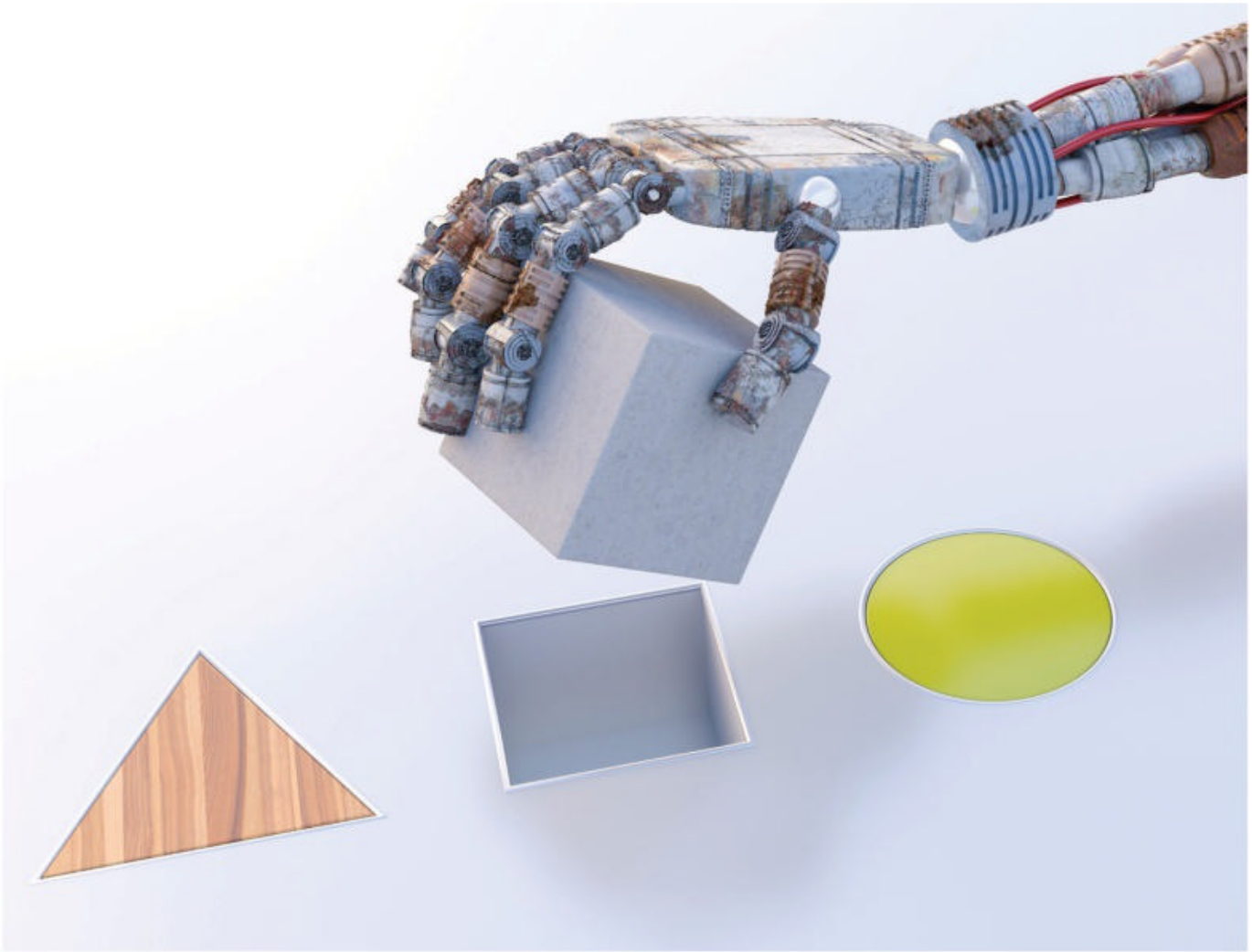
THE VIRTUAL MANAGER COLLECTION EBOOK + TOOLS

PRODUCT #10122E

store.hbr.org

1-800-988-0886 OR +1-617-783-7500

Quick Takes



WESTEND61/GETTY IMAGES

1. How to Choose Your First AI Project

→ by ANDREW NG

AI IS POISED to transform every industry, just as electricity did 100 years ago. It will create $13 trillion of GDP growth by 2030, according to McKinsey, most of which will be in noninternet sectors including manufacturing, agriculture, energy, logistics, and education. The rise of AI presents an opportunity for executives in every industry to differentiate and defend their businesses. But implementing a company-wide AI strategy is challenging, especially for legacy enterprises.

My advice for executives, in any industry, is to start small. The first step to building an AI strategy, drawn from the "AI Transformation Playbook," is to choose one or two company-level pilot AI projects. These projects will help your company gain momentum and firsthand knowledge of what it takes to build an AI product.

Five Traits of a Strong AI Pilot Project

Tapping the power of AI technologies requires customizing them to your business context. The purpose of your pilot projects is only partly to create value; more important, the success of these first projects will help convince stakeholders to invest in building up your company's AI capabilities.

When you're considering a pilot AI project, ask yourself the following questions:

Does the project give you a quick win? Use your first AI pilot project to get the flywheel turning as soon as possible. Choose a project that can be done quickly (ideally within six to 12 months) and has a high chance of success. Instead of doing only one pilot project, do two or three to increase the odds of creating at least one significant success.

Is the project either too trivial or too unwieldy in size? Your pilot project does not have to be the most valuable AI application as long as it delivers a quick win. But it should be meaningful enough that a success convinces other company leaders to invest in further AI projects.

In the early days of leading the Google Brain team, I faced widespread skepticism within Google about the potential of deep learning. Speech recognition was much less important to

Your pilot project should be meaningful enough that a success convinces other company leaders to invest in further AI projects.

Google than web search and advertising, so I had my team take on speech as our first internal customer. By helping the speech team build a much more accurate recognition system, we convinced other teams to have faith in Google Brain. For our second project, we worked with Google Maps to increase data quality. Each successful project increased the momentum in the flywheel, and Google Brain played a leading role in turning Google into the great AI company it is today.

Is your project specific to your industry? By choosing a company-specific project, your internal stakeholders can directly understand the value. For example, if you run a medical devices company, building an AI+Recruiting project to automatically screen résumés is a bad idea for two reasons: (1) There's a high chance someone else will build an AI+Recruiting platform that serves a much larger user base and will outperform what you could do in-house and/or undercut you. And (2) this project is less likely to convince the rest of your company that AI is worth investing in than if your pilot project applied AI to medical devices. It is more valuable to build a health-care-specific AI system—anything ranging from using AI to assist doctors with crafting treatment plans to streamlining the hospital check-in process through automation to offering personalized health advice.

Are you accelerating your pilot project with credible partners? If you are still building up your AI team, consider working with external partners to bring in AI expertise quickly. Eventually, you will want to have your own in-house AI team; however, waiting to build a team before executing might be too slow relative to the pace of AI's rise.

Is your project creating value? Most AI projects create AI value in one of three ways: reducing costs (automation creates opportunities for cost reduction in almost every industry), increasing revenue (recommendation and prediction systems increase sales and efficiency), or launching new lines of business (AI enables new projects that were not possible before).

You can create value even without having "big data," which is often overhyped. Some businesses, such as web search, have a long tail of queries, so search engines with more data do perform better. However, not all businesses have this amount of data, and it may be possible to build a valuable AI system with perhaps as few as 100 to 1,000 data records (though more does not hurt). Do not choose projects just because you have a lot of data in industry X and believe the AI team will figure out how to turn this data into value. Projects like this tend to fail. It is important to develop a thesis up front about how specifically an AI system will create value.

Setting Up Your AI Project for Success

So what do these traits look like in practice?

AI is automation on steroids. A rich source of ideas for AI projects will lie in automating tasks that humans are doing today using a technology called supervised learning. You will find that AI is good at automating *tasks*, rather than *jobs*. Try to identify the specific tasks that people are doing and examine if any can be automated. For example, the tasks involved in a radiologist's job may include reading X-rays, operating imaging machinery, consulting with colleagues, and planning for surgeries. Rather than trying to automate their entire job, consider if just one of the tasks could be automated or made faster through partial automation.

Before executing on an AI pilot, I recommend clearly stating the desired timeline and outcome, and allocating a reasonable budget to the team.

Appoint a leader. Choose someone who can work cross functionally and bridge both AI and your industry's domain experts. That way, when the project succeeds, it will influence the rest of the organization. Their goal is *not* to build an AI start-up; it is to build a successful project that will influence the rest of the company's beliefs and state of knowledge about AI as a first step toward building future projects.

Conduct business value *and* technical diligence. Make sure that, if the project is executed successfully, the business leaders agree that it will create sufficient value for the business. But also make sure that the project is feasible. Technical diligence can take a small number of weeks, requiring a technical team to examine what data you have and perhaps even carry out small-scale experiments.

Build a small team. I have seen numerous pilot ideas that were executed with about five to 15 people. The exact level of resources varies wildly per project, but scoping projects that can be done with a small team ensures that everyone can know everyone else and work cross functionally, and perhaps also makes the allocation of resources more painless. While some projects today require 100+ (or even 1,000+) engineers to do well, such a high level of resourcing is most likely not necessary for your pilot AI project.

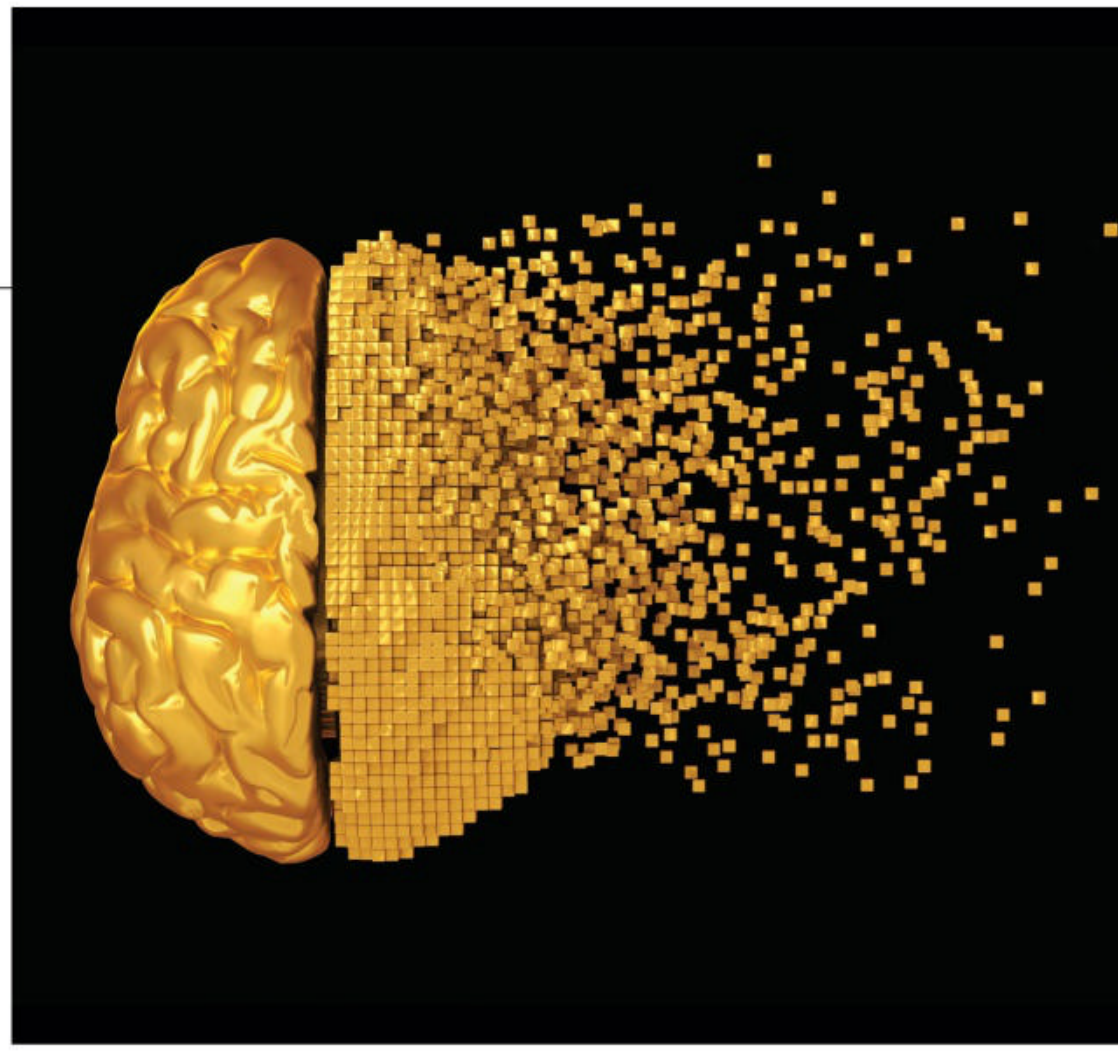
Communicate. When the pilot project hits key milestones, especially when it delivers a successful result, give the team an internal platform—ranging from talks, to awards, to even external PR—to allow their work to become known inside the company. Making sure the project team is recognized by the CEO and is visibly successful will be a key part to building momentum. If you have an AI technology team working with a business team, the business team should receive plenty of credit and rewards for the success. This will encourage other business teams to jump into AI as well.

Having led Google Brain and Baidu's AI Group, which were respectively leading forces for turning Google and Baidu into great AI companies, I think most companies can and should become good at AI. Your goal should not be to compete with the leading internet companies but to master AI for your vertical industry sector. And remember: The first step is to select the right pilot projects and execute on them.

HBR Reprint H04S3S

Originally published on HBR.org February 6, 2019

Andrew Ng *is the founder and CEO of Landing AI, the former chief scientist of Baidu, a cochairman and cofounder of Coursera, the former founder and lead of Google Brain, and an adjunct professor at Stanford University.*



2. Is Your Data Infrastructure Ready for AI?

→ by SETH EARLEY and JOSH BERNOFF

EVERY BIG COMPANY now manages a proliferation of sites, apps, and technology systems for interacting with buyers and managing everything in the business, from customers and clients to inventory and products. These systems are spitting out data continuously. But even after multiple generations of investments and billions of dollars of digital transformations, organizations struggle to use that data to improve customer service, reduce costs, and speed the core processes that provide competitive advantage.

AI was supposed to help with that. But as an executive at a major life insurance company recently told me (Seth), "Every one of our competitors and most of the organizations of our size in other industries have spent at least a few million dollars on failed AI initiatives." Why?

My 20 years of experience working with companies on their information technology

3DSCULPTOR/GETTY IMAGES

A company's ontology is the master knowledge scaffolding of the organization.

have shown me the reason: Promises of AI vendors don't pay off unless a company's data systems are properly prepared for AI. Data is locked in silos, inaccessible, poorly structured, and most important, not organized in such a way as to be used as the fuel that makes AI work. Instead, to reap the benefits of AI, companies need to create an ontology (a comprehensive characterization of the architecture of all its data).

You may have read advice that you should start small with AI initiatives. (I suggested that myself a few years back.) And it's true—the "AI Lite" approach can generate some quick wins. But as AI initiatives inevitably multiply throughout the organization, the limits of scattered experiments become more glaring. When you feed such AI programs with different types of (sometimes incompatible) data sources, the result will entangle you in complexity. Soon you'll have a slew of one-off AI pilots connected to your existing data systems in a way that fails to deliver broader, more strategic benefits for your business.

AI is now far enough along that a more cohesive approach is needed—a key that brings together all of your company's data. That's where the ontology comes in.

An ontology is a consistent representation of data and data relationships within your business, a model of all the elements that go into and connect your various information systems: the products and services, solutions and processes, organizational structures, protocols, customer characteristics, manufacturing methods, knowledge, content, and data of all types. It is the master knowledge scaffolding of the organization. Without a consistent, thoughtful approach to developing, applying, and evolving an ontology, AI systems can develop in only a piecemeal, fragmented way—they will lack the underpinning that would allow them to be smart enough to make an impact. The ontology is at the heart of the information design of the AI-powered enterprise, an investment that will continue to pay off as AI becomes more pervasive.

Building an Ontology at Applied Materials

Let's look at a B2B service firm, Applied Materials, and how they developed an ontology that showcases these benefits. (I advised them on this project.)

Applied Materials works with semiconductor manufacturers to fix problems that slow or halt production at semiconductor plants—problems that can cost millions of dollars a day until they're remedied. Because until recently the knowledge needed to keep a plant running was distributed among 14 different systems within Applied Materials, technicians spent up to 40% of their time searching for answers. Every plant is unique, so finding the right answer for a particular plant's problem is both essential and challenging. Techs tended to hedge their bets by stocking their service vehicles with a variety of costly components, tying up tens of millions of dollars in inventory.

The techs needed a consistent, efficient experience. But to build it, Applied Materials needed a way to organize the diverse sources of information available to the techs and integrate them into a single interface.

The ontology we created for the company included all the multiple vocabularies, relationships, and hierarchies in all the systems the techs used. It defined relationships for the short names that one system used to refer to a part and the stock number that another system used to refer to the same part. We parsed and classified troubleshooting documents with "text analytics," a method that learns from a model document, then extracts information from similar documents and makes all the knowledge accessible with a common language. When we were done, we had created a common language with a master set of relationships that connected the information in the company's customer management systems, service ticket tracking systems, databases of solutions, parts catalogs, and all the other systems.

Then, after we had created the ontology, we needed to actually apply it to all the underlying data, connect it to the techs' experience, and make it accessible from a common interface. After everything was tagged, Applied Materials incorporated the ontology in multiple systems and processes, applied it to existing documents, incorporated it into workflows for new documents and solutions, and connected it to ERP and digital asset management systems. This enabled us to create the new system we had dreamed of that empowered field techs to get exactly the information they needed, efficiently, from whatever sources were

most appropriate to the task at hand. The logical groups of content, or "mental models"—organized in ways that aligned with their way of thinking about the problems they were solving—helped them locate exactly the information that would be most useful in a given context, like troubleshooting guides.

The resulting system reduced the time technicians spent searching for information by half. Applied Materials estimated the value of that savings at tens of millions of dollars per year. It even enabled technicians to reduce their component-part inventories. AI was a part of the solution—but it depended on the underlying ontology to get appropriate access to the data it needed.

How to Build an Ontology

As you might imagine, creating an ontology that will make sense of all your company's data isn't an easy undertaking. Based on the many companies I've walked through the process, here are the key steps:

1. Identify data pain points. Audit where information problems and bottlenecks are impairing the function of the business, like the Applied Materials techs' frustrating solution searches. Look for root causes that are generating multiple problems. This helps people understand how solving the bigger data architecture problem can generate widespread benefits in the enterprise.

2. Generate solutions based on root causes. Start to imagine what a potential solution could look like—such as the more efficient search system Applied Materials so sorely needed. Begin with the tasks that will have the biggest impact on the business if made more efficient.

3. Understand use cases. Consider who will benefit from each solution—customer service staff, product developers, or maybe customers on an e-commerce site?—and exactly what tasks the solution will help them perform more efficiently. Develop a model for what the individual must do to achieve their desired outcome, then validate that by testing it out with actual workers doing their jobs. Learn how they describe the ideal answer, then match up how they describe it with the ways that other workers in other roles describe the same, or related, types of problems.

4. Set the ontology's organizing principles. This is the key to the ontology. With the knowledge about how people are using the information, you can begin to architect the details of how to organize and categorize the data you have to make it part of the solutions you are creating. For example, at Applied Materials, master concepts had names like "Account," "User," "Business Unit," and "Solution" that matched up to a variety of terms within the company's legacy systems—and to the mental models that the techs had for what was going on in any part of the plant.

Once you've created an ontology this way (and a system to maintain it), the benefits multiply. It becomes an essential part of the solution not only to today's information problem but to the next problem that arises, and the next. Keep in mind that the ontology is a continuously evolving entity—as products, services, markets, competitors, and customer needs change, an intentional process for changing the ontology will keep it fresh and relevant.

It's increasingly clear that AI is going to be solving many of those problems in the future, whether that's by means of a customer service chatbot, a system that surfaces signals about business efficiencies and breakdowns, or any of a thousand other applications. The ontology, as a representation of what matters within the company and what makes it unique, unifies those solutions and makes them more than just a quick fix that will just as quickly become obsolete. If you suspect that AI is the future of business—a conclusion I'm certain of—then creating an ontology is an essential investment to prepare your enterprise to realize the benefits of that future. It's a concept that, managed and applied appropriately, makes the difference between the promise of AI and delivering sustainably on that promise.

HBR Reprint H05KGH

Originally published on HBR.org April 28, 2020

Seth Earley *is the founder and CEO of Earley Information Science and award-winning author of* The AI-Powered Enterprise: Harness the Power of Ontologies to Make Your Business Smarter, Faster, and More Profitable *(LifeTree, 2020).* ***Josh Bernoff*** *is the author or coauthor of six business strategy books, including* Writing Without Bullshit: Boost Your Career by Saying What You Mean *(Harper Business, 2016).*

Harvard Business Review

"Every time a woman in my company has asked me for advice, there is an episode I can cite."

Women at Work

BACK FOR A SEVENTH SEASON!

HBR's acclaimed podcast helps you go after what you want, get through tough times, take care of yourself, and shine.

Available on Apple Podcasts, Spotify, and wherever you listen to podcasts.

hbr.org/podcasts/women-at-work

CO-SPONSORED BY

HubSpot

HARVARD BUSINESS SCHOOL
Executive Education

ORIGINALLY PUBLISHED JULY–AUGUST 2018

Collaborative Intelligence: Humans and AI Are Joining Forces

→ by H. JAMES WILSON and PAUL R. DAUGHERTY

ARTIFICIAL INTELLIGENCE IS becoming good at many "human" jobs—diagnosing disease, translating languages, providing customer service—and it's improving fast. This is raising reasonable fears that AI will ultimately replace human workers throughout the economy. But that's not the inevitable, or even most likely, outcome. Never before have digital tools been so responsive to us, nor we to our tools. While AI will radically alter how work gets done and who does it, the technology's larger impact will be in complementing and augmenting human capabilities, not replacing them.

Certainly, many companies have used AI to automate processes, but those that deploy it mainly to displace employees will see only short-term productivity gains. In our research involving 1,500 companies, we found that firms achieve the most significant performance improvements when humans and machines work together (see the exhibit "The Value of Collaboration"). Through such collaborative intelligence,

Illustration by **ENISAURUS**

humans and AI actively enhance each other's complementary strengths: the leadership, teamwork, creativity, and social skills of the former, and the speed, scalability, and quantitative capabilities of the latter. What comes naturally to people (making a joke, for example) can be tricky for machines, and what's straightforward for machines (analyzing gigabytes of data) remains virtually impossible for humans. Business requires both kinds of capabilities.

To take full advantage of this collaboration, companies must understand how humans can most effectively augment machines, how machines can enhance what humans do best, and how to redesign business processes to support the partnership. Through our research and work in the field, we have developed guidelines to help companies achieve this and put the power of collaborative intelligence to work.

Humans Assisting Machines

Humans need to perform three crucial roles. They must *train* machines to perform certain tasks; *explain* the outcomes of those tasks, especially when the results are counterintuitive or controversial; and *sustain* the responsible use of machines (by, for example, preventing robots from harming humans).

Training. Machine-learning algorithms must be taught how to perform the work they're designed to do. In that effort, huge training data sets are amassed to teach machine-translation apps to handle idiomatic expressions, medical apps to detect disease, and recommendation engines to support financial decision making. In addition, AI systems must be trained how best to interact with humans. While organizations across sectors are now in the early stages of filling trainer roles, leading tech companies and research groups already have mature training staffs and expertise.

Consider Microsoft's AI assistant, Cortana. The bot required extensive training to develop just the right personality: confident, caring, and helpful but not bossy. Instilling those qualities took countless hours of attention by a team that included a poet, a novelist, and a playwright. Similarly, human trainers were needed to develop the personalities of Apple's Siri and Amazon's Alexa to ensure that they accurately reflected their companies' brands. Siri, for example, has just a touch of sassiness, as consumers might expect from Apple.

AI assistants are now being trained to display even more complex and subtle human traits, such as sympathy. The start-up Koko, an offshoot of the MIT Media Lab, has developed technology that can help AI assistants seem to commiserate. For instance, if a user is having a bad day, the Koko system doesn't reply with a canned response such as "I'm sorry to hear that." Instead it may ask for more information and then offer advice to help the person see his issues in a different light. If he were feeling stressed, for instance, Koko might recommend thinking of that tension as a positive emotion that could be channeled into action.

Explaining. As AIs increasingly reach conclusions through processes that are opaque (the so-called black-box problem), they require human experts in the field to explain their behavior to nonexpert users. These "explainers" are particularly important in evidence-based industries, such as law and medicine, where a practitioner needs to understand how an AI weighed inputs into, say, a sentencing or medical recommendation. Explainers are similarly important in helping insurers and law enforcement understand why an autonomous car took actions that led to an accident—or failed to avoid one. And explainers are becoming integral in regulated industries—indeed, in any consumer-facing industry where a machine's output could be challenged as unfair, illegal, or just plain wrong. For instance, the European Union's new General Data Protection Regulation (GDPR) gives consumers the right to receive an explanation for any algorithm-based decision, such as the rate offer on a credit card or mortgage. This is one area where AI will contribute to *increased* employment: Experts estimate that companies will have to create about 75,000 new jobs to administer the GDPR requirements.

Sustaining. In addition to having people who can explain AI outcomes, companies need "sustainers"—employees who continually work to ensure that AI systems are functioning properly, safely, and responsibly.

For example, an array of experts sometimes referred to as safety engineers focus on anticipating and trying to prevent harm by AIs. The developers of industrial robots that work alongside people have paid careful attention to ensuring that they recognize humans nearby and don't endanger them. These

Human-machine collaboration enables companies to interact with employees and customers in novel, more effective ways.

experts may also review analysis from explainers when AIs do cause harm, as when a self-driving car is involved in a fatal accident.

Other groups of sustainers make sure that AI systems uphold ethical norms. If an AI system for credit approval, for example, is found to be discriminating against people in certain groups (as has happened), these ethics managers are responsible for investigating and addressing the problem. Playing a similar role, data compliance officers try to ensure that the data that is feeding AI systems complies with the GDPR and other consumer-protection regulations. A related data-use role involves ensuring that AIs manage information responsibly. Like many tech companies, Apple uses AI to collect personal details about users as they engage with the company's devices and software. The aim is to improve the user experience, but unconstrained data gathering can compromise privacy, anger customers, and run afoul of the law. The company's "differential privacy team" works to make sure that while the AI seeks to learn as much as possible about a group of users in a statistical sense, it is protecting the privacy of individual users.

Machines Assisting Humans

Smart machines are helping humans expand their abilities in three ways. They can *amplify* our cognitive strengths; *interact* with customers and employees to free us for higher-level tasks; and *embody* human skills to extend our physical capabilities.

Amplifying. Artificial intelligence can boost our analytic and decision-making abilities by providing the right information at the right time. But it can also heighten creativity. Consider how Autodesk's Dreamcatcher AI enhances the imagination of even exceptional designers. A designer provides Dreamcatcher with criteria about the desired product—for example, a chair able to support up to 300 pounds, with a seat 18 inches off the ground, made of materials costing less than $75, and so on. She can also supply information about other chairs that she finds attractive. Dreamcatcher then churns out thousands of designs that match those criteria, often sparking ideas that the designer might not have initially considered. She can then guide the software, telling it which chairs she likes or doesn't, leading to a new round of designs.

Throughout the iterative process, Dreamcatcher performs the myriad calculations needed to ensure that each proposed design meets the specified criteria. This frees the designer to concentrate on deploying uniquely human strengths: professional judgment and aesthetic sensibilities.

Interacting. Human-machine collaboration enables companies to interact with employees and customers in novel, more effective ways. AI agents like Cortana, for example, can facilitate communications between people or on behalf of people, such as by transcribing a meeting and distributing a voice-searchable version to those who couldn't attend. Such applications are inherently scalable—a single chatbot, for instance, can provide routine customer service to large numbers of people simultaneously, wherever they may be.

Idea in Brief

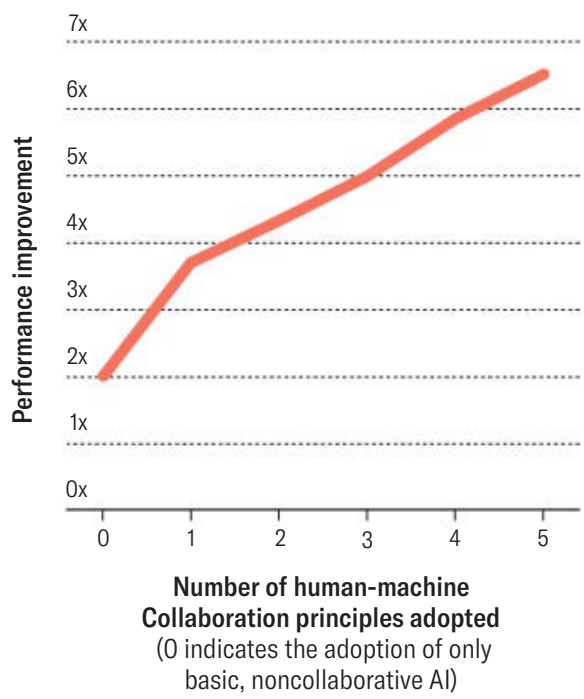
THE OUTLOOK
Artificial intelligence is transforming business—and having the most significant impact when it augments human workers instead of replacing them.

THE DETAILS
Companies see the biggest performance gains when humans and smart machines collaborate. People are needed to train machines, explain their outputs, and ensure their responsible use. AI, in turn, can enhance humans' cognitive skills and creativity, free workers from low-level tasks, and extend their physical capabilities.

THE PRESCRIPTION
Companies should reimagine their business processes, focusing on using AI to achieve more operational flexibility or speed, greater scale, better decision making, or increased personalization of products and services.

The Value of Collaboration

Companies benefit from optimizing collaboration between humans and artificial intelligence. Five principles can help them do so: Reimagine business processes; embrace experimentation/employee involvement; actively direct AI strategy; responsibly collect data; and redesign work to incorporate AI and cultivate related employee skills. A survey of 1,075 companies in 12 industries found that the more of these principles companies adopted, the better their AI initiatives performed in terms of speed, cost savings, revenues, or other operational measures.



SEB, a major Swedish bank, now uses a virtual assistant called Aida to interact with millions of customers. Able to handle natural-language conversations, Aida has access to vast stores of data and can answer many frequently asked questions, such as how to open an account or make cross-border payments. She can also ask callers follow-up questions to solve their problems, and she's able to analyze a caller's tone of voice (frustrated versus appreciative, for instance) and use that information to provide better service later. Whenever the system can't resolve an issue—which happens in about 30% of cases—it turns the caller over to a human customer-service representative and then monitors that interaction to learn how to resolve similar problems in the future. With Aida handling basic requests, human reps can concentrate on addressing more-complex issues, especially those from unhappy callers who might require extra hand-holding.

Embodying. Many AIs, like Aida and Cortana, exist principally as digital entities, but in other applications the intelligence is embodied in a robot that augments a human worker. With their sophisticated sensors, motors, and actuators, AI-enabled machines can now recognize people and objects and work safely alongside humans in factories, warehouses, and laboratories.

In manufacturing, for example, robots are evolving from potentially dangerous and "dumb" industrial machines into smart, context-aware "cobots." A cobot arm might, for example, handle repetitive actions that require heavy lifting, while a person performs complementary tasks that require dexterity and human judgment, such as assembling a gear motor.

Hyundai is extending the cobot concept with exoskeletons. These wearable robotic devices, which adapt to the user and location in real time, will enable industrial workers to perform their jobs with superhuman endurance and strength.

Reimagining Your Business

In order to get the most value from AI, operations need to be redesigned. To do this, companies must first discover and describe an operational area that can be improved. It might be a balky internal process (such as HR's slowness to fill staff positions), or it could be a previously intractable problem that can now be addressed using AI (such as quickly identifying adverse drug reactions across patient populations). Moreover, a number of new AI and advanced analytic techniques can help surface previously invisible problems that are amenable to AI solutions (see the sidebar "Revealing Invisible Problems").

Next, companies must develop a solution through co-creation—having stakeholders envision how they might collaborate with AI systems to improve a process. Consider the case of a large agricultural company that wanted to deploy AI technology to help farmers. An enormous amount of data was available about soil properties, weather patterns, historical harvests, and so forth, and the initial plan was to build an AI application that would more accurately predict future crop yields. But in discussions with farmers, the company learned of a more pressing need. What farmers really wanted was a system that could provide real-time recommendations on how to increase productivity—which crops to plant, where to grow them, how much nitrogen to use in the soil, and so on. The company developed an AI system to provide such advice, and the initial outcomes were promising; farmers were happy about the crop yields obtained with the AI's guidance. Results

At Mercedes-Benz, cobot arms guided by human workers pick up and place heavy parts, becoming an extension of the worker's body.

from that initial test were then fed back into the system to refine the algorithms used. As with the discovery step, new AI and analytic techniques can assist in co-creation by suggesting novel approaches to improving processes.

The third step for companies is to scale and then sustain the proposed solution. SEB, for example, originally deployed a version of Aida internally to assist 15,000 bank employees but thereafter rolled out the chatbot to its one million customers.

Through our work with hundreds of companies, we have identified five characteristics of business processes that companies typically want to improve: flexibility, speed, scale, decision making, and personalization. When reimagining a business process, determine which of these characteristics is central to the desired transformation, how intelligent collaboration could be harnessed to address it, and what alignments and trade-offs with other process characteristics will be necessary.

Flexibility. For Mercedes-Benz executives, inflexible processes presented a growing challenge. Increasingly, the company's most profitable customers had been demanding individualized S-class sedans, but the automaker's assembly systems couldn't deliver the customization people wanted.

Traditionally, car manufacturing has been a rigid process with automated steps executed by "dumb" robots. To improve flexibility, Mercedes replaced some of those robots with AI-enabled cobots and redesigned its processes around human-machine collaborations. At the company's plant near Stuttgart, Germany, cobot arms guided by human workers pick up and place heavy parts, becoming an extension of the worker's body. This system puts the worker in control of the build of each car, doing less manual labor and more of a "piloting" job with the robot.

The company's human-machine teams can adapt on the fly. In the plant, the cobots can be reprogrammed easily with a tablet, allowing them to handle different tasks depending on changes in the workflow. Such agility has enabled the manufacturer to achieve unprecedented levels of customization. Mercedes can individualize vehicle production according to the real-time choices consumers make at dealerships, changing everything from a vehicle's dashboard components to the seat leather to the tire valve caps. As a result, no two cars rolling off the assembly line at the Stuttgart plant are the same.

Speed. For some business activities, the premium is on speed. One such operation is the detection of credit-card fraud. Companies have just seconds to determine whether they should approve a given transaction. If it's fraudulent, they will most likely have to eat that loss. But if they deny a legitimate transaction, they lose the fee from that purchase and anger the customer.

Like most major banks, HSBC has developed an AI-based solution that improves the speed and accuracy of fraud detection. The AI monitors and scores millions of transactions daily, using data on purchase location and customer behavior, IP addresses, and other information to identify subtle patterns that signal possible fraud. HSBC first implemented the system in the United States, significantly reducing the rate of undetected fraud and false positives, and then rolled it out in the UK and Asia. A different AI system used by Danske Bank improved its fraud-detection rate by 50% and decreased false positives by 60%. The reduction in the number of false positives frees investigators to concentrate their efforts on equivocal transactions the AI has flagged, where human judgment is needed.

The fight against financial fraud is like an arms race: Better detection leads to more-devious criminals, which leads to better detection, which continues the cycle. Thus the algorithms and scoring models for combating fraud have a very short shelf life and require continual updating. In addition, different countries and regions use different models. For these reasons, legions of data analysts, IT professionals, and experts in financial fraud are needed at the interface between humans and machines to keep the software a step ahead of the criminals.

Scale. For many business processes, poor scalability is the primary obstacle to improvement. That's particularly true of processes that depend on intensive human labor with minimal machine assistance. Consider, for instance, the employee recruitment process at Unilever. The consumer goods giant was looking for a way to diversify its 170,000-person workforce. HR determined that it needed to focus on entry-level hires and then fast-track the best into management. But the company's existing processes weren't able to evaluate potential recruits in sufficient numbers—while giving each applicant individual attention—to ensure a diverse population of exceptional talent.

Enhancing Performance

At organizations in all kinds of industries, humans and AI are collaborating to improve five elements of business processes.

FLEXIBILITY	Auto manufacturing	Mercedes-Benz	Assembly robots work safely alongside humans to customize cars in real time.
	Product design	Autodesk	Software suggests new product design concepts as a designer changes parameters such as materials, cost, and performance requirements.
	Software development	Gigster	AI helps analyze any type of software project, no matter the size or complexity, enabling humans to quickly estimate the work required, organize experts, and adapt workflows in real time.
SPEED	Fraud detection	HSBC	AI screens credit- and debit-card transactions to instantly approve legitimate ones while flagging questionable ones for humans to evaluate.
	Cancer treatment	Roche	AI aggregates patient data from disparate IT systems, speeding collaboration among specialists.
	Public safety	Singapore government	Video analytics during public events predicts crowd behavior, helping responders address security incidents rapidly.
SCALE	Recruiting	Unilever	Automated applicant screening dramatically expands the pool of qualified candidates for hiring managers to evaluate.
	Customer service	Virgin Trains	Bot responds to basic customer requests, doubling the volume handled and freeing humans to address more-complex issues.
	Casino management	GGH Morowitz	Computer-vision system helps humans continuously monitor every gaming table in a casino.
DECISION MAKING	Equipment maintenance	General Electric	"Digital twins" and Predix diagnostic application provide techs with tailored recommendations for machine maintenance.
	Financial services	Morgan Stanley	Robo-advisers offer clients a range of investment options based on real-time market information.
	Disease prediction	Icahn School of Medicine at Mount Sinai	Deep Patient system helps doctors predict patients' risk of specific disease, allowing preventive intervention.
PERSONALIZATION	Guest experience	Carnival Corporation	Wearable AI device streamlines the logistics of cruise-ship activities and anticipates guest preferences, facilitating tailored staff support.
	Health care	Pfizer	Wearable sensors for Parkinson's patients track symptoms 24/7, allowing customized treatment.
	Retail fashion	Stitch Fix	AI analyzes customer data to advise human stylists, who give customers individualized clothing and styling recommendations.

Here's how Unilever combined human and AI capabilities to scale individualized hiring: In the first round of the application process, candidates are asked to play online games that help assess traits such as risk aversion. These games have no right or wrong answers, but they help Unilever's AI figure out which individuals might be best suited for a particular position. In the next round, applicants are asked to submit a video in which they answer questions designed for the specific position they're interested in. Their responses are analyzed by an AI system that considers not just what they say but also their body language and tone. The best candidates from that round, as judged by the AI, are then invited to Unilever for in-person interviews, after which humans make the final hiring decisions.

It's too early to tell whether the new recruiting process has resulted in better employees. The company has been closely tracking the success of those hires, but more data is still needed. It is clear, however, that the new system has greatly broadened the scale of Unilever's recruiting. In part because job seekers can easily access the system by smartphone, the number of applicants doubled to 30,000 within a year, the number of universities represented surged from 840 to 2,600, and the socioeconomic diversity of new hires increased. Furthermore, the average time from application to hiring decision has dropped from four months to just four weeks, while the time that recruiters spend reviewing applications has fallen by 75%.

Decision making. By providing employees with tailored information and guidance, AI can help them reach better decisions. This can be especially valuable for workers in the trenches, where making the right call can have a huge impact on the bottom line.

Consider the way in which equipment maintenance is being improved with the use of "digital twins"—virtual models of physical equipment. General Electric builds such software models of its turbines and other industrial products and continually updates them with operating data streaming from the equipment. By collecting readings from large numbers of machines in the field, GE has amassed a wealth of information on normal and aberrant performance. Its Predix application, which uses machine-learning algorithms, can now predict when a specific part in an individual machine might fail.

This technology has fundamentally changed the decision-intensive process of maintaining industrial equipment. Predix might, for example, identify some unexpected rotor wear and tear in a turbine, check the turbine's operational history, report that the damage has increased fourfold over the past few months, and warn that if nothing is done, the rotor will lose an estimated 70% of its useful life. The system can then suggest appropriate actions, taking into account the machine's current condition, the operating environment, and aggregated data about similar damage and repairs to other machines. Along with its recommendations, Predix can generate information about their costs and financial benefits and provide a confidence level (say, 95%) for the assumptions used in its analysis.

Without Predix, workers would be lucky to catch the rotor damage on a routine maintenance check. It's possible that it would go undetected until the rotor failed, resulting in a costly shutdown. With Predix, maintenance workers are alerted to potential problems before they become serious, and they have the needed information at their fingertips to make good decisions—ones that can sometimes save GE millions of dollars.

Personalization. Providing customers with individually tailored brand experiences is the holy grail of marketing. With AI, such personalization can now be achieved with previously unimaginable precision and at vast scale. Think of the way the music streaming service Pandora uses AI algorithms to generate personalized playlists for each of its millions of users according to their preferences in songs, artists, and genres. Or consider Starbucks, which, with customers' permission, uses AI to recognize their mobile devices and call up their ordering history to help baristas make serving recommendations. The AI technology does what it does best, sifting through and processing copious amounts of data to recommend certain offerings or actions, and humans do what they do best, exercising their intuition and judgment to make a recommendation or select the best fit from a set of choices.

The Carnival Corporation is applying AI to personalize the cruise experience for millions of vacationers through a wearable device called the Ocean Medallion and a network that allows smart devices to connect. Machine learning dynamically processes

Revealing Invisible Problems

Former U.S. Defense Secretary Donald Rumsfeld once famously distinguished among "known knowns," "known unknowns," and "unknown unknowns"—things you're not even aware you don't know. Some companies are now using AI to uncover unknown unknowns in their businesses. Case in point: GNS Healthcare applies machine-learning software to find overlooked relationships among data in patients' health records and elsewhere. After identifying a relationship, the software churns out numerous hypotheses to explain it and then suggests which of those are the most likely. This approach enabled GNS to uncover a new drug interaction hidden in unstructured patient notes. CEO Colin Hill points out that this is not garden-variety data mining to find associations. "Our machine-learning platform is not just about seeing patterns and correlations in data," he says. "It's about actually discovering causal links."

the data flowing from the medallion and from sensors and systems throughout the ship to help guests get the most out of their vacations. The medallion streamlines the boarding and debarking processes, tracks the guests' activities, simplifies purchasing by connecting their credit cards to the device, and acts as a room key. It also connects to a system that anticipates guests' preferences, helping crew members deliver personalized service to each guest by suggesting tailored itineraries of activities and dining experiences.

The Need for New Roles and Talent

Reimagining a business process involves more than the implementation of AI technology; it also requires a significant commitment to developing employees with what we call "fusion skills"—those that enable them to work effectively at the human-machine interface. To start, people must learn to delegate tasks to the new technology, as when physicians trust computers to help read X-rays and MRIs. Employees should also know how to combine their distinctive human skills with those of a smart machine to get a better outcome than either could achieve alone, as in robot-assisted surgery. Workers must be able to teach intelligent agents new skills and undergo training to work well within AI-enhanced processes. For example, they must know how best to put questions to an AI agent to get the information they need. And there must be employees, like those on Apple's differential privacy team, who ensure that their companies' AI systems are used responsibly and not for illegal or unethical purposes.

We expect that in the future, company roles will be redesigned around the desired outcomes of reimagined processes, and corporations will increasingly be organized around different types of skills rather than around rigid job titles. AT&T has already begun that transition as it shifts from landline telephone services to mobile networks and starts to retrain 100,000 employees for new positions. As part of that effort, the company has completely overhauled its organizational chart: Approximately 2,000 job titles have been streamlined into a much smaller number of broad categories encompassing similar skills. Some of those skills are what one might expect (for example, proficiency in data science and data wrangling), while others are less obvious (for instance, the ability to use simple machine-learning tools to cross-sell services).

MOST ACTIVITIES AT the human-machine interface require people to *do new and different things* (such as train a chatbot) and to *do things differently* (use that chatbot to provide better customer service). So far, however, only a small number of the companies we've surveyed have begun to reimagine their business processes to optimize collaborative intelligence. But the lesson is clear: Organizations that use machines merely to displace workers through automation will miss the full potential of AI. Such a strategy is misguided from the get-go. Tomorrow's leaders will instead be those that embrace collaborative intelligence, transforming their operations, their markets, their industries, and—no less important—their workforces.

HBR Reprint R1804J

H. James Wilson *is global managing director of technology & business research at Accenture Research.* ***Paul R. Daugherty*** *is Accenture's group chief executive of technology and CTO. They are the coauthors of* Human + Machine: Reimagining Work in the Age of AI *(Harvard Business Review Press, 2018).*

Innovation Won't Wait for You

BY MICHAEL MIEBACH

MICHAEL MIEBACH IS THE CEO OF MASTERCARD

To read more insights from Mastercard's Become Index research with Harvard Business Review Analytic Services, visit:

BecomeIndex.com

The pandemic took many of our daily routines by storm—even chocolate.

Mars Inc., the maker of M&M's and Snickers, sells a lot of candy at checkout counters and during Halloween. But as with many other norms the pandemic changed, candy sales dropped significantly in 2020 as people avoided going to stores.

The company needed to find new opportunities to get in front of its customers and give them new ways to buy. So it recreated Halloween as a virtual event, letting customers buy digital credits that trick-or-treaters could redeem for real candy.

This response to the pandemic was representative of the intense creative disruption going on across industries. Businesses cast aside old models and quickly created new ones to take their place. Innovations that usually take years were rolled out in weeks.

As the world gathers its strength after the initial shock of the pandemic, there's now an opportunity for businesses to consider what they've learned through the peak of the crisis and where we all can—and should—act.

The most important insight: Innovation won't wait for you. Nor will consumers. In survey research for our Become Index, people said they would seek out companies and businesses that invest in the latest digital tools and prioritize data privacy and security. They're willing to pay more for a better experience—and willing to move away if they're let down. Many businesses are responding to these interests: We saw an increase of over 10% in executives putting a high priority on innovation, even at the height of the pandemic.

At the same time, many brands weren't meeting customers' demands last year. Our research found that plenty of companies missed opportunities either by failing to roll out new technologies their customers were looking for, like contactless payments; not making product returns easier; not building intuitive digital experiences that go beyond apps; or underestimating needed investments in data security.

There are useful instances of companies that focused on the right strategies. Take the example of European telco Vodafone and its accelerated investment in data networks. During the pandemic, Vodafone saw traffic on its networks spike 50% as people turned to e-commerce, remote work, and remote learning. While some companies might look to conserve cash in a crisis, Vodafone's customers needed more bandwidth, so the company built it out.

McDonald's also moved quickly to provide new digital tech that focused squarely on its customers. It invested in drive-through express lanes for mobile app users and GPS alerts for restaurant workers to prepare orders in time for a mobile-app customer's arrival.

At Mastercard, our work focused on safely moving billions of dollars in transactions that migrated to e-commerce and bringing touchless payments to stores around the world to give people more peace of mind.

These types of investments in digital tools and infrastructure were important when socially distanced services became a must. But they also increased choices and convenience for customers. Those types of innovations that serve a need in times of crisis and add long-term value are by far the most beneficial.

Of course, not every new concept will meet that high bar, but it's necessary to keep testing to uncover such innovations. That will help companies prepare for customers' changing needs and help them find their way through if a new storm hits.

Quick Takes

YAROSLAV KUSHTA/GETTY IMAGES

1. How Will AI Change Work? Here Are Five Schools of Thought

→ by MARK KNICKREHM

THE FUTURE OF the workforce is one of the biggest issues facing CEOs today. It's abundantly clear to all that AI, big data analytics, and advanced robotics make it possible for machines to take on tasks that once required a person to do them. How should companies prepare strategically to thrive in this world?

Views on what to expect vary dramatically. By some accounts, almost half of all jobs in the U.S. economy could be made obsolete. Others have described how intelligent machines will actually create jobs, including entirely new categories of jobs. Some people even talk about a world of superabundance where work will be about pursuing your passion, on your own terms.

It's critical for companies to understand the range of opinions on this issue, because implicitly or explicitly, they will influence the way business leaders create the workforce of the future. And while a lot will shake out in years to come, this issue is already front and center. Companies are making decisions today that will matter hugely to their ability to compete tomorrow and throughout the 2020s.

Most companies are already moving rapidly to

Business leaders must take steps now to shape their workforces for the emerging intelligent enterprise.

acquire new capabilities. In a new Accenture survey ("Reworking the Revolution," which was published on January 23, 2018) of 1,200 C-level executives worldwide, 75% say that they are currently accelerating investments in AI and other intelligent technologies. And 72% say they are responding to a competitive imperative—they recognize the need for new tools to keep up with rivals, both by improving productivity and by finding new sources of growth. Some companies are transforming themselves into "intelligent enterprises," in which all processes are digitized, decisions are data-driven, and machines do the heavy lifting—both physical and cognitive.

So, there's a great deal at stake in the debate over productivity and jobs. Leaders must understand the debate and be prepared to address tough questions: What kinds of new skills do we need? How should we be organized? How do we define jobs? How can we bring our people along with us in a way that benefits everyone?

Through research, we've identified five schools of thought in this debate.

1. **The Dystopians.** Man and machine will wage a Darwinian struggle that machines will win. AI systems will take on tasks at the heart of middle- and high-skill jobs, while robots will perform menial work that requires low-skill labor. The result will be massive unemployment, falling wages, and wrenching economic dislocation. Falling incomes will have grave consequences in places like the United States and Europe, where consumption accounts for 56% or 69% of GDP, respectively, requiring new social supports, such as a universal basic income.

2. **The Utopians.** Intelligent machines will take on even more work, but the result will be unprecedented wealth, not economic decline. AI and computing power will advance in the next two decades to achieve "the singularity"—when machines will be able to emulate the workings of the human brain in its entirety. Human brains will be "scanned" and "downloaded" to computers, and billions of replicated human brains will do most of the cognitive work, while robots will do all the heavy lifting. Economic output could double every three months. The singularity may even lead to a world where little human labor is required, a universal income program covers basic needs, and people apply their talents to meaningful pursuits.

3. **The Technology Optimists.** A burst of productivity has already begun but is not captured in official data because companies are still learning how intelligent technologies can change how they operate. When companies do take full advantage of intelligent technologies, a leap in productivity will produce a digital bounty—creating both economic growth and improvements in living standards not counted in GDP, such as consumer surplus (from better, cheaper products) and the value of free apps and information. However, on the basis of current trends, the bounty won't be distributed evenly, and many jobs will be displaced. To avoid negative income and employment effects, investments in education and training alongside investments in technology will be necessary.

4. **The Productivity Skeptics.** Despite the power of intelligent technologies, any gains in national productivity levels will be low. Combine that with headwinds from aging populations, income inequality, and the costs of dealing with climate change, and the United States will have near-zero GDP growth. In the end, there isn't much to do except brace for stagnant growth in advanced economies.

5. **The Optimistic Realists.** Digitization and intelligent machines can spur productivity gains that match previous technology waves. Productivity will advance rapidly in certain sectors and for high-performing companies. New jobs will be created, but intelligent technologies may exacerbate the trends of the recent past, in which demand rose for both high- and low-skill workers whose jobs could be easily automated, while demand for middle-skill workers fell. With no simple solutions, more research is needed into the true relationship between productivity, employment, and wages to uncover effective responses.

Three Actions for Shaping the Future

Our crystal ball for what things might look like in 10 years is cloudy. What we do know is that business leaders must take steps now to shape their workforces for the

emerging intelligent enterprise. Three imperatives are critical:

Use technology to augment human skills and reinvent operating models. Companies that think beyond labor substitution and cost savings will see a much greater payoff. For example, a new class of adaptive robots can function safely alongside workers and take on difficult and tedious work. Consider this example: At BMW's plant in Spartanburg, South Carolina, robots are installing door-sealing gaskets, an awkward and tiring job for workers. This speeds up the line, improves quality, and gives workers more time to do higher-value work. Researchers estimate that using adaptive robots this way could cut time wasted on non-value-added work by 25%. Employee surveys show that workers have more-positive views of the new robots, which they regard as useful helpers. Away from the factory, companies are using AI to off-load routine work from employees and give them new analytical tools to improve customer experience and discover new possibilities for products, services, and business models that drive growth.

Take the opportunity to redefine jobs and rethink organizational design. Companies cannot optimize their investments if they have the same old job descriptions and organizational structures. Executives should assess the tasks that need to be done, anticipate which ones will be transferred to machines, then reconfigure jobs by adding new tasks or creating entirely different roles that are needed for managing intelligent technologies. A factory worker, for example, can be trained to run robots. AI systems also need human help to train and correct algorithms and override fallible machine judgment. For example, at Stitch Fix, an online clothing subscription service, 3,400 human stylists work with an AI recommendation engine to make personalized suggestions for customers. The machines give stylists the speed they need to be productive, and the stylists provide the additional judgment needed for accurate recommendations (and fewer returns). To function effectively, an intelligent enterprise should have a nonhierarchical organization in which employees collaborate across functional and operational silos. This enables the intelligent enterprise to act quickly on the insights from data-crunching machines and deploy human talent to swarm on problems, experiment, iterate, and get solutions into the market.

Make employees your partners in building the intelligent enterprise. To strike the right balance between investing in intelligent technologies and maintaining existing businesses, companies need help from their employees. In our (above referenced) research, we have found that employees are far more willing—even eager—to master new technologies than employers appreciate. They want to learn new skills, not least because they know they will need them to remain employed. Investments in both technology and training will help companies make a smooth transition to the intelligent enterprise. Companies that do this stand to outperform competitors because they will unleash the human talents that machines still can't match and that are essential to growth—creativity, empathy, communications, adaptability, and problem-solving. "As basic automation and machine learning move toward becoming commodities," says Devin Fidler, research director at the Institute for the Future, "uniquely human skills will become more valuable."

The debate over technology and jobs will rage on. Business leaders must follow this debate—and participate in it, too. And much more research is needed to fully understand the implications of intelligent technologies on work. In the meantime, companies that actively seize control of what can be done to prepare will position themselves to thrive in this exciting new era.

HBR Reprint H0449F

Originally published on HBR.org January 24, 2018

The author thanks his colleagues in Accenture Research—Svenja Falk, David Light, and Geoffrey Lewis—for their contributions to this article.

***Mark Knickrehm** was the group chief executive for Accenture Strategy until he retired in 2020.*

ANDRZEJ WOJCICKI/GETTY IMAGES

2. AI Should Augment Human Intelligence, Not Replace It

→ by DAVID DE CREMER and GARRY KASPAROV

IN AN ECONOMY where data is changing how companies create value—and compete—experts predict that using AI at a larger scale will add as much as $15.7 trillion to the global economy by 2030. As AI is changing how companies work, many believe that who does this work will change, too—and that organizations will begin to replace human employees with intelligent machines. This is already happening: Intelligent systems are displacing humans in manufacturing, service delivery, recruitment, and the financial industry, consequently moving human workers toward lower-paid jobs or making them unemployed. This trend has led some to conclude that in 2040 our workforce may be totally unrecognizable.

Are humans and machines really in competition with each other though? The history of work—particularly since the Industrial Revolution—is the history of people outsourcing their labor to machines. While that began with rote, repetitive physical tasks like weaving, machines have evolved to the point where they can now do what we might think of as complex cognitive work, such as math equations, language and speech recognition, and writing. Machines thus seem ready to replicate the work of our minds, not just our bodies. In the 21st century, AI is evolving to be superior to humans in many tasks, and we seem ready to outsource our intelligence to technology. With this latest trend, it appears as if there's nothing that can't soon be automated, meaning that no job is safe from being off-loaded to machines.

This vision of the future of work has taken the shape of a zero-sum game, in which there can only be one winner.

We believe, however, that this view of the role AI will play in the workplace is wrong. The question of whether AI will replace human workers assumes that AI and humans have the same qualities and abilities—but, in reality, they don't. AI-based machines are fast, more accurate, and consistently rational, but they aren't intuitive, emotional, or

culturally sensitive. And, it's exactly these abilities that humans possess and that make us effective.

Machine Intelligence vs. Human Intelligence

In general, people recognize today's advanced computers as intelligent because they have the potential to learn and make decisions on the basis of the information they take in. But while we may recognize that ability, we possess a decidedly different type of intelligence.

In its simplest form, AI is a computer acting and deciding in ways that seem intelligent. In line with Alan Turing's philosophy, AI imitates how humans act, feel, speak, and decide. This type of intelligence is extremely useful in an organizational setting: Because of its imitating abilities, AI has the quality to identify informational patterns that optimize trends relevant to the job. In addition, contrary to humans, AI never gets physically tired and as long it's fed data, it will keep going.

These qualities mean that AI is perfectly suited for lower-level routine tasks that are repetitive and take place within a closed management system. In such a system, the rules of the game are clear and not influenced by external forces. Think, for example, of an assembly line where workers are not interrupted by external demands and influences like work meetings. As a case in point, the assembly line is exactly the place where Amazon placed algorithms in the role of managers to supervise human workers and even fire them. As the work is repetitive and subject to rigid procedures optimizing efficiency and productivity, AI can perform more accurately than human supervisors.

Human abilities, however, are more expansive. Contrary to AI abilities, which are responsive only to the data available, humans can imagine, anticipate, feel, and judge changing situations, which allows them to shift from short-term to long-term concerns. These abilities are unique to humans and do not require a steady flow of externally provided data to work, as is the case with AI.

In this way humans represent what we call authentic intelligence—a different type of AI, if you will. This type of intelligence is needed when open systems are in place. In an open management system, the team or organization is interacting with the external environment and therefore has to deal with outside influences. This work setting requires the ability to anticipate and work with, for example, sudden changes and distorted information exchange, while at the same time be creative in distilling a vision and future strategy. In open systems, transformation efforts are continuously at work and effective management of that process requires authentic intelligence.

Although artificial intelligence (referred to as AI1 here) seems to be the opposite of authentic intelligence (referred to as AI2 here), they are also complementary. In the context of organizations, both types of intelligence offer a range of specific talents.

Which talents—operationalized as abilities needed to meet performance requirements—are necessary to perform best? It is, first of all, important to emphasize that talent can win games, but often it will not win championships—teams win championships. For this reason, we believe that the combination of the talents included in both AI1 and AI2, working in tandem, will make for the future of intelligent work. It will create the kind of intelligence that will allow organizations to be more efficient and accurate, but at the same time also creative and proactive. This other type of AI we call augmented intelligence (referred to as AI3 here).

The Third Type of AI: Augmented Intelligence

What will AI3 be able to offer that AI1 and AI2 can't? The second author of this article has some unique insight here: He is known for winning chess championships while at the same time has the distinctive experience of being the first human to lose a high-level game to a machine. In 1997, chess grand master Garry Kasparov lost a game to an IBM supercomputer program called Deep Blue. That left him to rethink how the intellectual game of chess could be approached differently, not simply as an individual effort but as a collaborative one. And, with the unexpected victory of Deep Blue, he decided to try collaborating with an AI.

In a match in 1998 in León, Spain, Kasparov partnered with a PC running the chess software of his choice—an arrangement called "advanced chess"—in a match against the Bulgarian Veselin Topalov, whom he had beaten 4–0 a month earlier. This time, with both players supported by computers, the match ended in a 3–3 draw. The use of a PC appeared to nullify the calculative and

Greater productivity and the automation of cognitively routine work is a boon, not a threat.

strategic advances Kasparov usually displayed over his opponent.

The match provided an important illustration of how humans might work with AI. Afterward, Kasparov noted that using a PC allowed him to focus more on strategic planning while the machine took care of the calculations. Nevertheless, he also stressed that simply putting together the best human player and the best PC did not, in his eyes, reveal games that were perfect. As with human teams, the power of working with AI comes from how the person and computer complement each other; the best players and the most powerful AIs partnering up doesn't necessarily produce the best results.

Once again, the chess world offers a useful test case for how this collaboration can play out. In 2005 the online chess-playing site Playchess.com hosted what it called a "freestyle" chess tournament in which anyone could compete in teams with other players or computers. What made this competition interesting is that several groups of grand masters working with computers also participated in this tournament. Predictably, most people expected that one of these grand masters in combination with a supercomputer would dominate the competition—but that's not what happened. The tournament was won by a pair of amateur American chess players using three computers. It was their ability to coordinate and coach their computers effectively that defeated the combination of a smart grand master and a PC with great computational power.

This surprising result underscores an important lesson: The process of how players and computers interact determines how efficient the partnership will be. Or, as Kasparov expressed it, "Weak human + machine + better process was superior to a strong computer alone and, more remarkably, superior to a strong human + machine + inferior process."

Recommendations

The enhancing and collaborative potential that we envision stands in stark contrast to the zero-sum predictions of what AI will do to our society and organizations. Instead, we believe that greater productivity and the automation of cognitively routine work is a boon, not a threat. After all, new technology always has disruptive effects early on in the implementation and development phases and reveals its real value only after some time.

This reality, however, does not mean that we have to wait patiently until this value eventually reveals itself—very much the opposite! Our principal challenge as businesspeople is to anticipate what AI means in relationship to how humans think and act, and to integrate the new technologies ambitiously and strategically into our organizations. We can't just passively wait for it to overtake traditional methods. So, what can we do at this moment to ensure the integration of the different AIs to make our organizations work effectively?

First, teams will gradually become composed of humans and nonhumans working together, which we refer to as the "new diversity." The psychology of the new diversity will bring with it the risk that stereotypical beliefs and biases can easily influence decisions and teamwork. Machine as a nonhuman coworker may be met with distrust and negative expectations as with any other out-group member and as such may encourage humans to share less information and avoid working with the machine. Team leaders will need to respond to such negative team dynamics and be trained so that they understand the reality of those negative beliefs and the consequences.

Second, the new shape of teams will call for leaders who are skilled in bringing different parties together. In the future, creating inclusive teams by aligning man and machine will be an important ability for people to be trained in and develop. As the above examples show, to achieve better performance by employing these new diversity teams, leaders will need to transform themselves into masters of coordinating and coaching team processes.

Third, team processes will need to be managed effectively by a human. For humans to align the strengths and weaknesses of man and machine, they will need to be educated to understand how AI works and what it can be used for, and decide—by means of the judgment abilities of their authentic intelligence—how it can be used best to foster performance serving human interests.

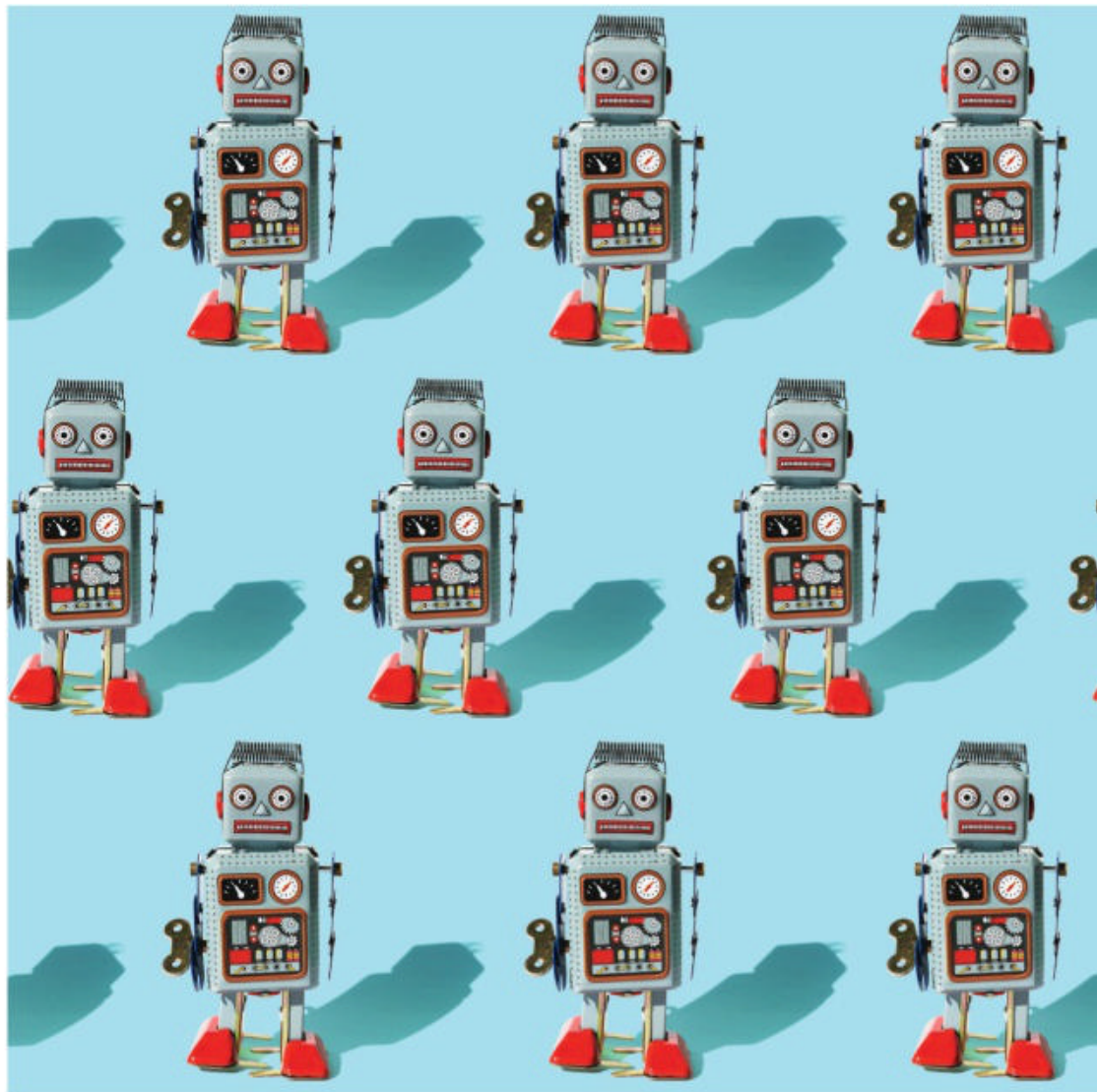
Augmented intelligence, as the third type of AI, is the step forward to the future of intelligent work. The future of work is a concept used to indicate the growth of employees and their performance in more-efficient

ways. The debate on this topic, however, has become quite ambiguous in its intentions. Specifically, because of cost-cutting strategies, businesses today are in a stage where machines are often introduced as the new superemployee that may ultimately leave humans in an inferior role to serve machines. An essential element of a truly intelligent type of future of work, however, means expanding the workforce of which both humans and machine will be a part, but with the aim to improve humanity and well-being while also being more efficient in the execution of our jobs. Augmented intelligence is indeed collaborative in nature, but a collaborative effort in service of humans.

HBR Reprint H068OV

Originally published on HBR.org March 18, 2021

David De Cremer *is a provost's chair and professor in management and organizations at NUS Business School, National University of Singapore, where he is also the founder and director of the Centre on AI Technology for Humankind. Before moving to NUS, he was the KPMG endowed chaired professor in management studies and current honorary fellow at Judge Business School, University of Cambridge. He has been named one of the world's top 30 management gurus and speakers and one of the 30 global thought leaders to watch by Thinkers50. His most recent book is* Leadership by Algorithm: Who Leads and Who Follows in the AI Era? *(Harriman House, 2020), which was named as one of the 15 summer books to read by Wharton.* ***Garry Kasparov*** *is the chairman of the Human Rights Foundation and founder of the Renew Democracy Initiative. He writes and speaks frequently on politics, decision-making, and human–machine collaboration. He became the youngest world chess champion in history at 22 in 1985 and retained the top rating in the world for 20 years. His famous matches against the IBM supercomputer Deep Blue in 1996 and 1997 were key to bringing AI, and chess, into the mainstream. His latest book is* Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins *(PublicAffairs, 2017).*



WESTEND61/GETTY IMAGES

3. Why Robots Won't Steal Your Job

→ by NAHIA ORDUÑA

SCIENCE FICTION FILMS and novels usually portray robots as one of two things: destroyers of the human race or friendly helpers. The common theme is that these stories happen in an alternate universe or a fantasy version of the future. Not here, and not now—until recently. The big difference is that the robots have come not to destroy our lives but to disrupt our work.

Last year, the World Economic Forum released a report estimating that by 2025, 85 million jobs may be

displaced by a shift in labor division between humans and machines.

For new grads entering the workforce, or young professionals looking to land their first jobs, this is news worth paying attention to. Entry-level positions that include routine tasks are precisely the ones disappearing.

What exactly is happening?

Thanks to advances in technology, some computers can conduct business processes without our margin of error. Natural language processing (NLP) allows chatbots to understand speech and provide technical support to customers in a variety of industries, including food and retail services. HR departments and finance companies use robotic process automation (RPA) to verify payroll systems, create email reports, and manage expenses, among other tasks typically handled by employees. And computer vision now makes it possible for machines to scan barcodes and track packages without the help of human hands.

You may be thinking that it's finally happened: The robots have won.

But don't panic just yet. The same report that predicts robots will soon steal our jobs also says that even more roles will open as a result of this shift—97 million to be exact. These are the "jobs of the future," and they are actually better opportunities, specifically for early-career professionals.

There are two reasons:

- The more computers are trained to conduct high-repetitive tasks that are often assigned to entry-level employees, the more roles focused on complex tasks with competitive salaries will arise in their place. This means that young professionals may have a wider range of interesting careers to choose from.
- People just entering the workforce usually struggle to land roles with higher salaries because they have to compete with senior candidates. This competitive disadvantage disappears as new types of roles—roles that *no one* has done before—are created. Younger workers are less likely to be forced to compete with their seniors, and more likely to be pioneers.

That said, if you are just entering the workforce, you may feel overwhelmed by the plethora of the new (and sometimes confusing) opportunities. To figure out which path you want to take first, you'll need to spend some

The Top 20 Roles in Increasing and Decreasing Demand Across Industries

According to the World Economic Forum Future of Jobs Report 2020, 85 million jobs may be displaced by the shift in labor between humans and machines by 2025, while 97 million new roles may emerge.

	Increasing Demand	Decreasing Demand
1	Data analysts and scientists	Data entry clerks
2	AI and machine-learning specialists	Administrative and executive secretaries
3	Big data specialists	Accounting, bookkeeping, and payroll clerks
4	Digital-marketing and -strategy specialists	Accountants and auditors
5	Process automation specialists	Assembly and factory workers
6	Business development professionals	Business services and administration managers
7	Digital transformation specialists	Client information and customer service workers
8	Information security analysts	General and operations managers
9	Software and applications developers	Mechanics and machinery repairers
10	Internet of Things (IoT) specialists	Material-recording and stock-keeping clerks
11	Project managers	Financial analysts
12	Business services and administration managers	Postal service clerks
13	Database and network professionals	Sales reps: wholesale and manufacturing, technical, and scientific products
14	Robotics engineers	Relationship managers
15	Strategic advisers	Bank tellers and related clerks
16	Management and organization analysts	Door-to-door sales, news, and street vendors
17	Fintech engineers	Electronics and telecoms installers and repairers
18	Mechanics and machinery repairers	Human resources specialists
19	Organizational development specialists	Training and development specialists
20	Risk management specialists	Construction laborers

Source: Future of Jobs Report 2020, World Economic Forum

time researching what areas feel right for you.

Ask yourself these three questions to identify the "jobs of tomorrow" worth applying to and to future-proof your career.

1. What are the new roles in my area of interest? Innovations in technology are driving the creation of new jobs. To figure out what roles you may want to pursue, you first need to understand what is happening in your area of interest. My first piece of advice is pretty intuitive: Do a simple Google search. Include the name of your industry plus key phrases like "future of work" or "job trends in [industry]." For example, you might search for "HR future of work," "digital marketing trends," or "banking in 2025."

If you need a little inspiration before you start your research, refer to the graphic on the next page, which highlights jobs in various sectors that are expected to boom in the next few years.

Once you learn what's available in your field and decide what roles sound most exciting to you, you can refine your search on sites like Indeed, LinkedIn, and Glassdoor.

For example, did you study marketing? Your search may teach you that companies are hiring new roles, like digital marketing specialist—a position tasked with making data-driven decisions to ensure campaigns are successful—that can be filled by employees at any level. Is your passion HR? Scoring the internet will show you that information technology recruiter and HR analyst are two jobs expected to pop up in the next few years. Both roles are responsible for collecting and analyzing people data to help companies make better choices.

On LinkedIn, I'd recommend researching people who may already be in the roles that interest you. Follow them, react to their posts, and join the virtual events where they are presenting—but don't just message them out of the blue. Reach out directly only when you can clearly share how their work or ideas have impacted you. From my experience, people are more likely to respond when you show you have done your homework.

In your message, you can explain that, as someone who is just entering the workforce, you're curious to learn more about their journey. If you reach out to enough people, a few will respond, especially if you are authentic and approach them with a personalized note.

Ask to schedule a virtual coffee and use that time to learn more about what their job entails, how they came into the position, and what skills you may need to sharpen to stand out as a candidate for a similar role. This will give you a better idea of what companies are hiring for right now and the level of experience you need to be considered for those positions.

The connections you make may even end up getting you a referral down the line.

2. Which new skills do I need to master and how? Most people graduate college with a solid foundation in the area of their degree. Decades ago, this was enough to get you hired. But today, that foundation is only a starting point.

Technologies are moving so fast that to be a competitive candidate, you will need to learn the latest tools and digital trends in your industry.

For example, you may have studied marketing in college, but by the time you graduate, new social media channels and publishing tools will have emerged. You may have gotten your law degree, but the same year you apply to your first job, you might learn that law firms have started digitizing contracts and are now prioritizing entry-level roles to help their clients maintain data privacy or conduct internet transactions. You may have graduated with a major in economics, but were digital trading and cryptocurrency management covered in your coursework?

This is where upskilling comes in. Upskilling is all about gaining knowledge to accelerate your career by closing skill gaps that can help you advance in your industry. If you're the marketing major, it's about mastering those new social media channels. If you're the law student, it's about studying data protection. And if you're the economics wiz, it's about educating yourself on digital trading trends.

How do you start? List the skills related to the role (or roles) you are most interested in applying to. You can find these within job descriptions or through the connections you make and research you conduct on sites like LinkedIn.

What skills are required for your dream role that you don't yet have? What trainings or certifications

The Jobs of Tomorrow

New technologies, products, and services are driving greater demand for new jobs in the industries listed below. The roles are organized below their corresponding industries and are ordered by the scale of each opportunity.

CARE ECONOMY

Cloud Computing

1. Site reliability engineer
2. Platform engineer
3. Cloud engineer
4. DevOps engineer
5. Cloud consultant
6. DevOps manager

Content Production

1. Social media assistant
2. Social media coordinator
3. Content specialist
4. Content producer
5. Content writer
6. Creative copywriter

Data and AI

1. Artificial intelligence specialist
2. Data scientist
3. Data engineer
4. Big data developer
5. Data analyst
6. Analytics specialist
7. Data consultant
8. Insights analyst
9. Business intelligence developer
10. Analytics consultant

Engineering

1. Python developer
2. Full stack engineer
3. JavaScript developer
4. Back-end developer
5. Front-end engineer
6. NET software developer
7. Development specialist
8. Technology analyst

GREEN ECONOMY

Marketing

1. Growth hacker
2. Growth manager
3. Digital marketing specialist
4. Digital specialist
5. E-commerce specialist
6. Commerce manager
7. Head of digital
8. Digital marketing consultant
9. Digital marketing manager
10. Chief marketing officer

People and Culture

1. Information technology recruiter
2. Human resources partner
3. Talent acquisition specialist
4. Business partner
5. Human resources business partner

Product Development

1. Product owner
2. Quality assurance tester
3. Agile coach
4. Software quality assurance engineer
5. Product analyst
6. Quality assurance engineer
7. Scrum master
8. Digital product manager
9. Delivery lead

Sales

1. Customer success specialist
2. Sales development representative
3. Commercial sales representative
4. Business development representative
5. Customer specialist
6. Partnerships specialist
7. Chief commercial officer
8. Head of partnerships
9. Enterprise account executive
10. Business development specialist
11. Chief strategy officer
12. Head of business development

Note: Due to constraints related to data availability, emerging roles in the Care and Green Jobs industries are not listed.
Source: Future of Jobs Report 2020, World Economic Forum; LinkedIn Economic Graph

have other professionals in these positions acquired? For example, if you are searching for digital marketing positions, you may find that many open roles require "hands-on experience with Google Analytics" or "expertise in paid search." Create a matrix to prioritize each skill you want to acquire, ranking how useful it will be for your target jobs and how much time it would require to learn. A basic understanding of most skills won't take more than two to three weeks.

Many digital tools can be learned online for free through YouTube tutorials. You can master even more if you take online courses. Platforms like Coursera, Udacity, FutureLearn, Udemy, or LinkedIn Learning are great places to start, and some tech companies, including Amazon, offer free training too.

Don't forget to add your newly acquired skills to your résumé. Online course certificates prove that you are not only up-to-date on the latest trends but also willing to keep learning and adapt yourself to the new times.

3. What else can you offer? What makes you unique? It could be that you are fluent in multiple languages, you have exceptional creative writing skills, or you are a master networker, or even that you grew up in a different country and therefore have a deeper understanding of international markets or cultures.

Throughout my career, I have seen junior colleagues given more responsibilities faster for using the qualities that set them apart to

get ahead on their teams or in their industries. Ask yourself: What skills have I acquired over the years because of who I am, where I am from, or what I am passionate about? Maybe you organized several clubs throughout your college years and have exceptional organization and leadership skills. Or you've developed an exceptional work ethic through your background as an athlete.

If you are not sure what unique qualities you have to offer (sometimes it can be hard to take a step back and analyze ourselves in this way), email your fellow students, colleagues, friends, or family members asking what differentiates you from other people in their eyes. Many times, others see our strengths before we do.

The ability to combine the new tools you learn through upskilling with what makes you unique will help you build a résumé that stands out. And if you can provide specific examples of situations in which you used this combination of skills to accomplish a task or goal, it will give you a leg up during screenings and interviews.

Machines may be disrupting the entry-jobs market, but as a young professional, remember that technology is here to help you. Most companies worth applying to are looking for different people, with different interests and experiences, who can add new perspectives to their teams. Robots may have a smaller margin of technical error, but they lack your humanity, your quirks, and the rare and distinctive qualities that make you who you are. Blend those qualities with the technical skills required for the exciting roles coming your way, and you will be setting yourself up for success both now and in the future.

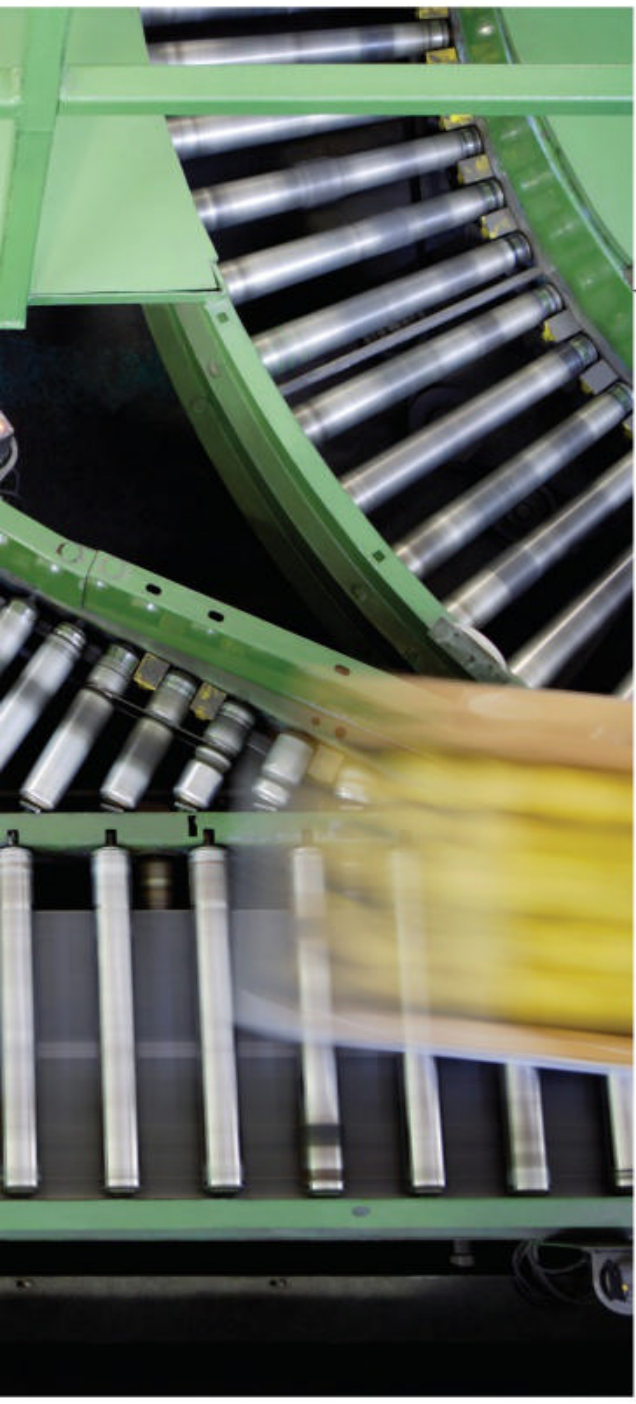
Originally published on Ascend, HBR's site for young professionals March 19, 2021.

Nahia Orduña *is an engineer and a technical leader at Amazon Web Services. She is the author of* Your Digital Reinvention: A Practical Guide for Discovering New Job Opportunities and Finding YOUR Place in the Future of Work *(independently published, 2020).*

4. How Amazon Automated Work and Put Its People to Better Use

→ by ALEX KANTROWITZ

AT AN AUTOMATION conference in late 2018, a high-ranking banking official looked up from his buffet plate and stated his objective without hesitation: *I'm here,* he told me, *to eliminate full-time employees.* I was at the conference because after spending months researching how Amazon automates work at its headquarters,



PAUL TAYLOR/GETTY IMAGES

I was eager to learn what other firms thought about this powerful technology. After one short interaction, it was clear that some have it completely wrong.

For the past decade, Amazon has been pushing to automate office work under a program now known as Hands Off the Wheel. The purpose was not to eliminate jobs but to automate tasks so that the company could reassign people to build new products—to do more with the people on staff rather than doing the same with fewer people. The strategy appears to have paid off: At a time when it's possible to start new businesses faster and cheaper than ever before, Hands Off the Wheel has kept Amazon operating nimbly, propelled it ahead of its competitors, and has shown that automating in order to fire staff can mean missing big opportunities. As companies look at how to integrate increasingly powerful AI capabilities into their businesses, they'd do well to consider this example.

The animating idea behind Hands Off the Wheel originated at Amazon's South Lake Union office towers, where the company began automating work in the mid-2010s under an initiative some called Project Yoda. At the time, employees in Amazon's retail management division spent their days making deals and working out product promotions, as well as determining what items to stock in its warehouses, in what quantities, and for what price. But with two decades' worth of retail data at its disposal, Amazon's leadership decided to use "the force" (machine learning) to handle the formulaic processes involved in keeping warehouses stocked. "When you have actions that can be predicted over and over again, you don't need people doing them," Neil Ackerman, an ex-Amazon general manager, said.

The project began in 2012, when Amazon hired Ralf Herbrich as its director of machine learning and made the automation effort one of his launch projects. Getting the software to be good at inventory management and pricing predictions took years, Herbrich said, because his team had to account for low-volume product orders that befuddled its data-hungry machine-learning algorithms. By 2015, the team's machine-learning predictions were good enough that Amazon's leadership placed them in employees' software tools, turning them into a kind of copilot for human workers. But at that point the humans could override the suggestions, and many did, setting back progress.

Eventually, though, automation took hold. "It took a few years to slowly roll it out, because there was training to be done," Herbrich said. If the system couldn't make its own decisions, he explained, it couldn't learn. Leadership required employees to automate a large number of tasks, though that varied across divisions. "In 2016 my goals for Hands Off the Wheel were 80% of all my activity," one ex-employee said. By 2018 Hands Off the Wheel was part of business as usual. Having delivered on his project, Herbrich left the company in 2020.

The transition to Hands Off the Wheel wasn't easy. The retail-division employees were despondent at first, recognizing that their jobs were transforming. "It was a total change," the former employee mentioned above said. "Something that you were incentivized to do, now you're being disincentivized to do." Yet in time, many saw the logic. "When we heard that ordering was going to be automated by algorithms, on the one hand, it's like, 'OK, what's happening to my job?'" another former employee, Elaine Kwon, said. "On the other hand, you're also not surprised. You're like, 'OK, as a business this makes sense.'"

Although some companies might have seen an opportunity to reduce head count, Amazon assigned the employees new work. The company's retail-division workers largely moved into product and program manager jobs—fast-growing roles within Amazon that typically belong to professional

inventors. Product managers oversee new product development, while program managers oversee groups of projects. "People who were doing these mundane repeated tasks are now being freed up to do tasks that are about invention," Jeff Wilke, Amazon's departing CEO of Worldwide Consumer, said, "the things that are harder for machines to do."

Had Amazon eliminated those jobs, it would have made its flagship business more profitable but most likely would have caused itself to miss its next new businesses. Instead of automating to milk a single asset, it set out to build new ones. Consider Amazon Go, the company's checkout-free convenience store. Go was founded, in part, by Dilip Kumar, an executive once in charge of the company's pricing and promotions operations. While Kumar spent two years acting as a technical adviser to CEO Jeff Bezos, Amazon's machine-learning engineers began automating work in his old division, so he took a new lead role in a project aimed at eliminating the most annoying part of shopping in real life: checking out. Kumar helped dream up Go, which is now a pillar of Amazon's broader strategy.

If Amazon is any indication, businesses that reassign employees after automating their work will thrive. Those that don't risk falling behind. In shaky economic times, the need for cost-cutting could make it tempting to replace people with machines, but I'll offer a word of warning: Think twice before doing that. It's a message I wish I had shared with the banker.

HBR Reprint H05V26

Originally published on HBR.org September 16, 2020

Alex Kantrowitz *is the author of* Always Day One: How the Tech Titans Plan to Stay on Top Forever *(Portfolio, 2020), a new book about the tech giants' work culture. He writes about Amazon, Apple, Facebook, Google, and Microsoft in his weekly newsletter,* Big Technology.

ONURDONGEL/GETTY IMAGES

5. What We Often Get Wrong About Automation

→ by RAVIN JESUTHASAN and JOHN BOUDREAU

Many times automation results in higher compensation for workers, with those costs often being offset by greater productivity.

WHEN LEADERS DESCRIBE how advances in automation will affect job prospects for humans, predictions typically fall into one of two camps. Optimists say that machines will free human workers to do higher-value, more-creative work. Pessimists predict massive unemployment or, if they have a flair for the dramatic, a doomsday scenario in which humans' only job is to serve our robot overlords.

What almost everyone gets wrong is focusing exclusively on the idea of automation "replacing" humans. Simply asking which humans will be replaced fails to account for how work and automation will evolve. Our new book, *Reinventing Jobs: A 4-Step Approach for Applying Automation to Work,* argues that while automation can sometimes substitute for human work, more important, it also has the potential to create new, more-valuable, and more-fulfilling roles for humans.

The intense emphasis on automation eliminating certain jobs for humans often stems from a fixation on near-term economic outcomes, particularly a focus on a traditional "lift and shift" equation that measures success according to the reduction in labor costs resulting from having fewer workers. While most organizations also look for gains like improvements in speed, quality, and service, the focus is still on economic outcomes. Such a narrow focus overlooks some of the most important and vital benefits of optimizing work automation. Indeed, we have found that many times automation results in higher compensation for workers, with those costs often being offset by greater productivity. However, the benefits go beyond strict economic accounting. By reinventing jobs to optimize work between humans and automation, organizations can attract a larger and more qualified applicant pool and achieve better retention, greater safety, and increased diversity.

The reinvention of work in the oil and gas industry provides an excellent example of the more nuanced approach to optimized work automation, as well as the array of benefits that lie beyond the more obvious economics of reduced labor costs.

Work Reinvention on the Oil and Gas Rig

Traditionally, oil rig workers must be on-site, doing hands-on manual labor that is often dangerous. However, rapid advances in technology now make it possible to create a completely autonomous rig. The traditional economically motivated way to envision the benefits of this technology is to imagine how many workers doing today's oil rig jobs could be removed from the operation if the rigs were fully automated.

In contrast, the CEO and leadership team of one organization that worked with Ravin's firm, Willis Towers Watson, adopted a more deliberate and inventive strategy. First, they refocused their strategy to provide a more comprehensive extraction solution, envisioning the rig as a platform for services beyond simply drilling a hole in the ground, with automation driving that transformation.

Second, instead of fixating on eliminating human labor, they set out to optimize their human talent, shifting employees away from repetitive, physical, isolated, and dangerous work and toward more variable, mental, interactive, and less hazardous work. They reinvented their rig crews to provide many of the services that before had been provided by third parties, offering a complete oil field management solution.

Third, they recognized the fundamental value of deconstructing the existing jobs into work activities to more clearly reveal where automation would eliminate some work, augment other work, and create entirely new work that was previously not possible. The reinvented work activities fell into several categories:

- Activities centralized (sensors and AI monitor operations and transmit data to human workers in a control center)
- Activities shifted to other roles (maintenance previously done by teams on each rig are shifted to shared-service teams)
- Activities augmented (AI assists with monitoring rig performance, and sensors allow more-precise directional drilling)
- Activities eliminated (robotics now perform pipe running and other dangerous and dirty work)
- Activities created (previously lesser-skilled rig workers can now perform some electrical and mechanical engineering work)

These newly reinvented jobs required increasing pay levels by between 7% and 15%. A traditionalist aiming to cut labor costs with automation might see that as a losing proposition, but a more complete analysis reveals the benefits that easily offset the increased labor cost.

The Increased Profitability Payoff

Converting the rig into a platform for multiple oil field solutions and optimizing the core drilling performance produced a 45% increase in profitability and a significant reduction in performance variance between rigs. This substantial increase in overall rig profitability easily offset the increased labor cost. That profitability gain was partly due to envisioning the rigs and their workers more strategically—as part of a whole oil field solution rather than simply a drilling service for hire—and optimizing the role and impact of human labor. These results are impressive, but this strategic approach to work automation also produced two additional, unexpected gains.

Better Workforce Retention/Attraction

The reinvented jobs required very different skills. For example, the job of a "derrickman" was reinvented into a new job of "technician"; this shift eliminated the need for many basic manual skills but added several new competencies, such as advanced electrical and mechanical knowledge, and enhanced communication and collaboration skills.

As we noted, acquiring and training workers with these new skills required increasing pay levels. The increased pay levels helped address the significant industry challenge of attracting and retaining skilled talent. With the reinvented jobs, this company could better compete for top talent, not only with higher pay but also with more-desirable work, because many dirty and dangerous tasks were automated.

The organization expects to see more highly skilled job applicants apply, join, and stay.

The Diversity Dividend

The other surprising benefit from automation was an increase in workforce diversity. The traditional rig work was physically demanding, isolated, and dangerous. It attracted and retained workers who were overwhelmingly male, with family situations that allowed them to spend large blocks of time on the rig. After the organization reinvented this work, many activities that formerly were performed almost exclusively by men stationed on rigs in remote locations were now performed in a centralized control room in a major metropolitan area. The applicant and worker pool now included more women, and those with more-diverse family situations who possessed the cognitive skills required to perform this reinvented work. The diversity and inclusion of genders, family structures, regions, and ethnic backgrounds increased so much that for the first time the organization achieved its long-sought diversity goals.

Beyond the Organization

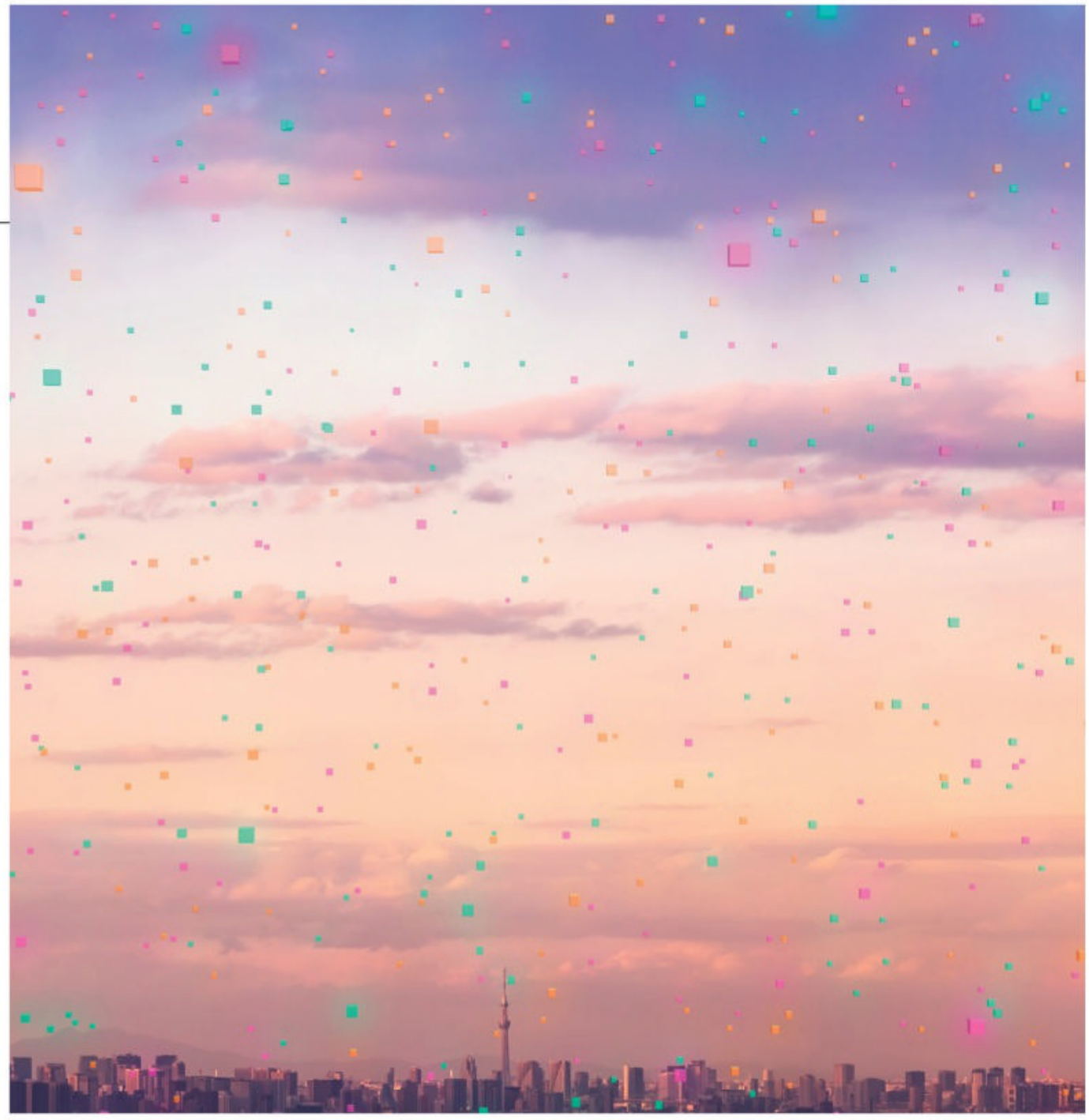
Debates about the impact of automation will rage on for the foreseeable future, but this example illustrates the many positive outcomes available through the careful and collaborative reinvention of jobs. This reinvention is not only vital to optimizing work and automation within a single firm; it is also important for empowering organizations to participate in the social conversation about future job creation and work. Political, social, economic, and regulatory debate typically relies on concepts like "good jobs," "lost jobs," and "preparing a workforce for the jobs of the future." By adopting a

systematic approach to work automation, organizations can better shape these debates by more clearly illuminating the full array of benefits and costs, both within the organization and for society more broadly.

HBR Reprint H04L63

Originally published on HBR.org October 11, 2018

Ravin Jesuthasan *is the global leader of transformation at Mercer. He is a member of the World Economic Forum's Steering Committee on Work and Employment and has been recognized as one of the 25 most influential consultants in the world. He is a coauthor, with John Boudreau, of* Reinventing Jobs: A 4-Step Approach for Applying Automation to Work *(Harvard Business Review Press, 2018) and numerous articles on work, automation, and human capital. Follow him on Twitter @ravinjesuthasan.* ***John Boudreau*** *is a professor emeritus of management and organization at the University of Southern California's Marshall School of Business and the research director at the university's Center for Effective Organizations. He is an expert on the bridge between superior human capital, talent, and sustainable success and the author of many books and articles, including his latest book with Ravin Jesuthasan,* Reinventing Jobs: A 4-Step Approach for Applying Automation to Work. *Follow him on Twitter @JohnWBoudreau.*



HIROSHI WATANABE/GETTY IMAGES

6. How to Harness the Digital Transformation of the Covid Era

→ by FEDERICA SALIOLA and ASIF M. ISLAM

DIGITAL TECHNOLOGY IS at the center of today's economic development debate due to its wide use during the Covid-19 outbreak. While there is no doubt that the pandemic is amplifying the adoption of new technologies, technological advancements have actually been changing the world for the past two decades, from living standards to the very nature of our work.

Fears of robot-induced unemployment are on the rise as tasks traditionally performed by humans are increasingly performed using robots and AI. The declining cost of machines threatens low-skill jobs and routine tasks—those most susceptible to automation and offshoring. Indeed, the number of robots operating worldwide is rising rapidly:

By the end of 2020, 3 million new industrial robots will be in operation, more than double the operational stock over the seven years spanning 2014–2020.

But technology *can* be a job creator, too. Increases in efficiency brought about by digital technology can help businesses expand. Digital platforms can create entirely new occupations and jobs. Businesses can reach out to remote markets lacking infrastructure. To reshape technology as a job creator, it's important to understand what, exactly, the current wave of technology is changing, and how policy makers and businesses can adapt to it.

The Foundations of Change

Even before the pandemic struck, some features of the state of technological progress were especially salient.

First, technology was already disrupting production processes, especially through the rapid scale-up of digital platforms. Digital technology has been challenging the traditional boundaries of firms, changing global value chains and the geography of jobs. After all, technology decreases the costs of doing business, complementing investments in infrastructure, free trade agreements, and other liberalization efforts to reduce trade barriers, which in turn expands global value chains and changes the geography of jobs. New business models—digital platform firms—have been able to evolve rapidly from local start-ups to global behemoths, often with few employees or tangible assets. Digital platforms have enabled clusters of businesses to form in underdeveloped rural areas.

Second, technology created seismic shifts in the mix of skills required to succeed in the labor market. While returns to routine, job-specific skills are declining, the premium for skills that cannot be replaced by robots has been increasing. These include cognitive skills such as critical thinking, as well as socio-behavioral skills such as managing and recognizing emotions that enhance teamwork. Earnings are higher for those who have a combination of these skills. The evolving world of work demands adaptable skills that enable workers to transfer more easily from one task to another. Since 2001, the share of employment in occupations intensive in nonroutine cognitive and socio-behavioral skills has increased from 19% to 23% in emerging economies, and from 33% to 41% in advanced economies.

Third, digital technology changed the terms of work. Rather than "standard" long-term contracts, digital technologies have given rise to more short-term work, often via online work platforms. These "gigs" make certain kinds of work more accessible and flexible. That said, despite the hype, the gig economy as of now has been slow to take over traditional occupations. The largest three global gig platforms—Freelancer of Australia, Upwork in the United States, and Zhubajia in China—have 60 million total users; only 0.3–0.5% of the active labor force participate in the gig economy globally.

How Will These Changes Play Out in the Post-Covid-19 World?

The pandemic will most likely reinforce these preexisting trends and increase the urgency of corresponding policy responses. Some points already seem clear. "Platform firms" are dominating markets even more. We are already seeing Amazon and Alibaba getting even bigger and stronger as brick-and-mortar stores are unable to compete. Companies will invest more in their ability to conduct business over the internet to be more resilient to potential lockdowns. Some "gig" jobs will also continue to grow. Firms may have more incentive to invest in automation and reshore production to shield against value chain disruption. Many businesses relying on imported inputs are facing lack of intermediate goods as value chains are disrupted. They may need to ensure that supplies are less vulnerable to travel restrictions.

Digital technology is also improving people's ability to work from home, although the amenability to remote work—which depends on the types of jobs and tasks to be done, as well as digital capacity—varies significantly across and within countries. Jobs that are conducive to remote work are more prevalent in rich countries, among workers with higher levels of education, and in salaried full-time jobs. Women and young workers are significantly less likely to work remotely. Digital infrastructure is scarce or of low quality in many developing countries.

Policies to Prepare for the Future of Work

The rapid spread of technology accelerated by the pandemic has led to a pressing need for businesses

The rapid spread of technology accelerated by the pandemic has led to a pressing need for businesses and governments to adapt.

and governments to adapt. Many businesses, especially in developing economies, are digitally disconnected. They may not have access to workers with the right skills and face challenging business environments. Workers, on the other hand, have little protection and do not have the skills or flexibility in labor markets to adapt. To face these challenges, businesses need to embrace technology and upgrade training programs to equip their workers with the best skills. Businesses can also consider apprenticeships to equip future workers with the right sets of skills.

Governments have an arsenal of policy options at their disposal, from incentives and regulations to infrastructure projects and taxation. Key priorities should be: (1) ramping up investment in human capital (knowledge, skills, and health) and lifelong learning if workers are to adapt to future labor markets; (2) strengthening social protections, expanding safety net coverage, and reforming financing arrangements and labor market norms to facilitate work transitions and reduce disincentives to the creation of formal jobs; (3) ensuring affordable access to the internet while adapting regulations to confront the challenges posed by digital platforms (such as data privacy and protection and competition rules); and (4) upgrading taxation systems to address tax avoidance and create fiscal space for universal social protection and human capital development.

Technology can be a boon to society if businesses and governments prepare and adapt. The pandemic has pushed societies to an inflection point where embracing technology is no longer an option but a necessity. It has also made workers more vulnerable. With the right steps and actions, businesses and governments can take the crisis as an opportunity to build for the future.

HBR Reprint H05VZU

Originally published on HBR.org September 24, 2020

Federica Saliola *is a lead economist in the Jobs Group of the World Bank and codirector of the* World Development Report 2019 The Changing Nature of Work, *which investigates the impact of technology on labor markets.* ***Asif M. Islam*** *is a senior economist for the Middle East and North Africa Region of the World Bank Group. His research focuses on private-sector development.*

ORIGINALLY PUBLISHED SEPTEMBER–OCTOBER 2019

Learning to Work with Intelligent Machines

Companies are deploying AI and robotics in ways that disrupt traditional training techniques such as mentoring and on-the-job learning. We need to understand how to combine the old with the new.

→ by MATT BEANE

IT'S 6:30 IN the morning, and Kristen is wheeling her prostate patient into the OR. She's a senior resident, a surgeon in training. Today she's hoping to do some of the procedure's delicate, nerve-sparing dissection herself. The attending physician is by her side, and their four hands are mostly in the patient, with Kristen leading the way under his watchful guidance. The work goes smoothly, the attending backs away, and Kristen closes the patient by 8:15, with a junior resident looking over her shoulder. She lets him do the final line of sutures. She feels great: The patient's going to be fine, and she's a better surgeon than she was at 6:30.

Fast-forward six months. It's 6:30 AM again, and Kristen is wheeling another patient into the OR, but this time for robotic prostate surgery. The attending leads the setup of a thousand-pound robot, attaching each of its four arms to the patient. Then he and Kristen take their places at a control console 15 feet away. Their backs are to the patient, and Kristen just watches as the attending remotely manipulates the robot's arms,

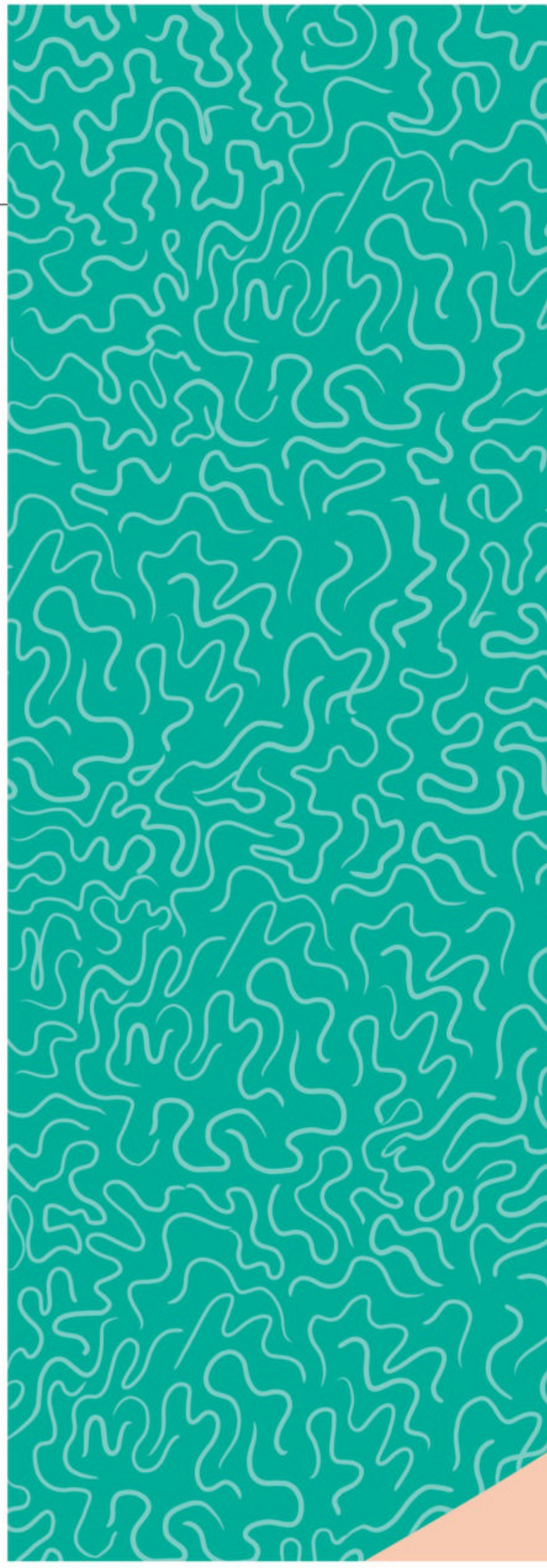
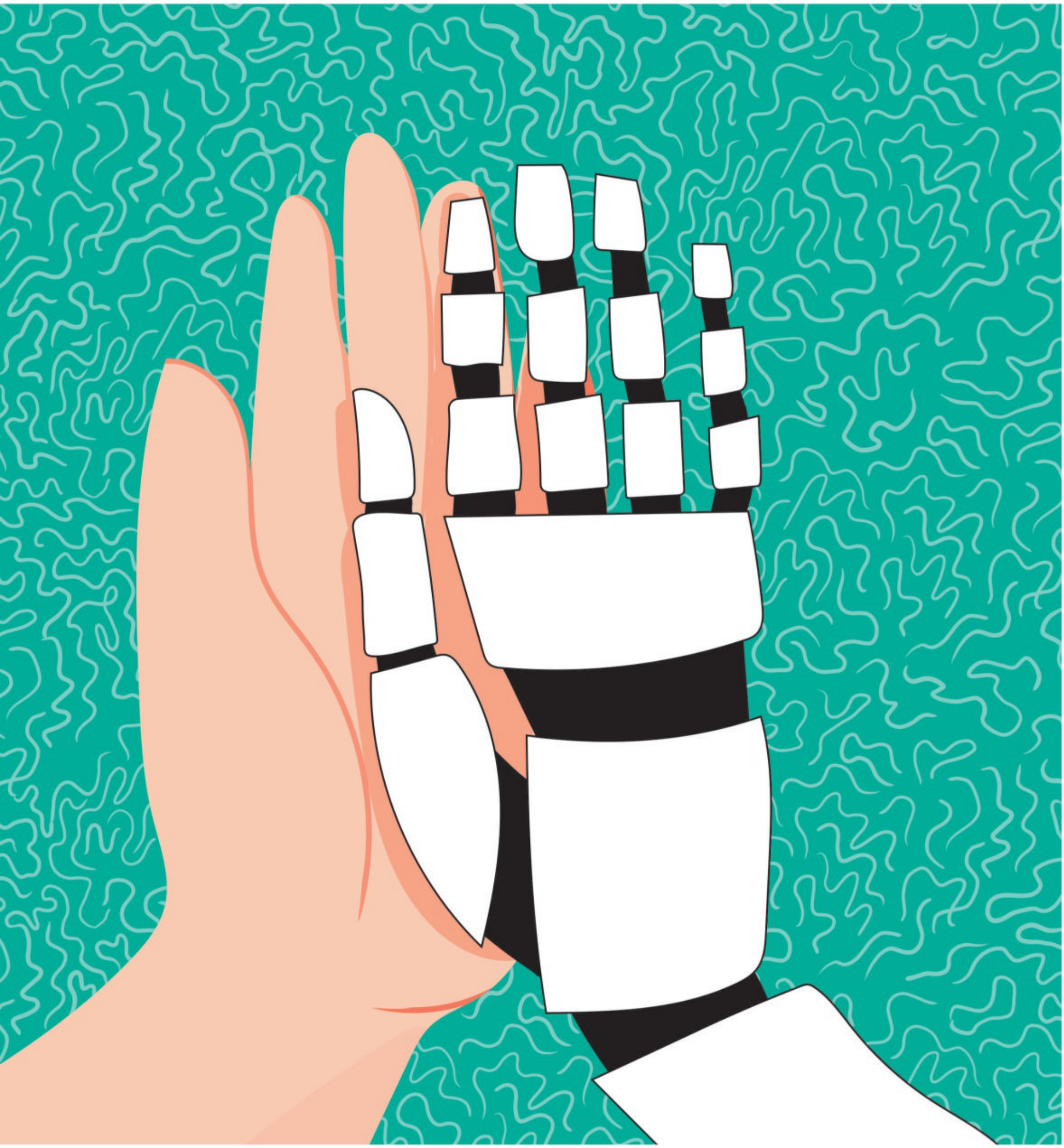


Illustration by **MICHELLE D'URBANO**



delicately retracting and dissecting tissue. Using the robot, he can do the entire procedure himself, and he largely does. He knows Kristen needs practice, but he also knows she'd be slower and would make more mistakes. So she'll be lucky if she operates more than 15 minutes during the four-hour surgery. And she knows that if she slips up, he'll tap a touch screen and resume control, very publicly banishing her to watch from the sidelines.

Surgery may be extreme work, but until recently surgeons in training learned their profession the same way most of us learned how to do our jobs: We watched an expert, got involved in the easier work first, and then progressed to harder, often riskier tasks under close supervision until we became experts ourselves. This process goes by lots of names: apprenticeship, mentorship, on-the-job learning (OJL). In surgery it's called *See one, do one, teach one.*

Critical as it is, companies tend to take on-the-job learning for granted; it's almost never formally funded or managed, and little of the estimated $366 billion companies spent globally on formal training in 2018 directly addressed it. Yet decades of research show that although employer-provided training is important, the lion's share of the skills needed to reliably perform a specific job can be learned only by doing it. Most organizations depend heavily on OJL: A 2011 Accenture survey, the most recent of its kind and scale, revealed that only one in five workers had learned any new job skills through formal training in the previous five years.

Today OJL is under threat. The headlong introduction of sophisticated analytics, AI, and robotics into many aspects of work is fundamentally disrupting this time-honored and effective approach. Tens of thousands of people will lose or gain jobs every year as those technologies automate work, and hundreds of millions will have to learn new skills and ways of working. Yet broad evidence demonstrates that companies' deployment of intelligent machines often blocks this critical learning pathway: My colleagues and I have found that it moves trainees away from learning opportunities and experts away from the action, and overloads both with a mandate to master old and new methods simultaneously.

How, then, will employees learn to work alongside these machines? Early indications come from observing learners engaged in norm-challenging practices that are pursued out of the limelight and tolerated for the results they produce. I call this widespread and informal process *shadow learning.*

Obstacles to Learning

My discovery of shadow learning came from two years of watching surgeons and surgical residents at 18 top-rated teaching hospitals in the United States. I studied learning and training in two settings: traditional ("open") surgery and robotic surgery. I gathered data on the challenges robotic surgery presented to senior surgeons, residents, nurses, and scrub technicians (who prep patients, help glove and gown surgeons, pass instruments, and so on), focusing particularly on the few residents who found new, rule-breaking ways to learn. Although this research concentrated on surgery, my broader purpose was to identify learning and training dynamics that would show up in many kinds of work with intelligent machines.

To this end, I connected with a small but growing group of field researchers who are studying how people work with smart machines in settings such as internet start-ups, policing organizations, investment banking, and online education. Their work reveals dynamics like those I observed in surgical training. Drawing on their disparate lines of research, I've identified four widespread obstacles to acquiring needed skills. Those obstacles drive shadow learning.

1. Trainees are being moved away from their "learning edge." Training people in any kind of work can incur costs and decrease quality, because novices move slowly and make mistakes. As organizations introduce intelligent machines, they often manage this by reducing trainees' participation in the risky and complex portions of the work, as Kristen found. Thus trainees are being kept from situations in which they struggle near the boundaries of their capabilities and recover from mistakes with limited help—a requirement for learning new skills.

The same phenomenon can be seen in investment banking. New York University's Callen Anthony found that junior analysts in one firm were increasingly being separated from senior partners as those partners interpreted algorithm-assisted company valuations in M&As. The junior analysts were tasked with simply pulling raw reports

Companies' deployment of intelligent machines often moves trainees away from learning opportunities and experts away from the action.

from systems that scraped the web for financial data on companies of interest and passing them to the senior partners for analysis. The implicit rationale for this division of labor? First, reduce the risk that junior people would make mistakes in doing sophisticated work close to the customer; and second, maximize senior partners' efficiency: The less time they needed to explain the work to junior staffers, the more they could focus on their higher-level analysis. This provided some short-term gains in efficiency, but it moved junior analysts away from challenging, complex work, making it harder for them to learn the entire valuation process and diminishing the firm's future capability.

2. Experts are being distanced from the work. Sometimes intelligent machines get between trainees and the job, and other times they're deployed in a way that prevents experts from doing important hands-on work. In robotic surgery, surgeons don't see the patient's body or the robot for most of the procedure, so they can't directly assess and manage critical parts of it. For example, in traditional surgery, the surgeon would be acutely aware of how devices and instruments impinged on the patient's body and would adjust accordingly; but in robotic surgery, if a robot's arm hits a patient's head or a scrub is about to swap a robotic instrument, the surgeon won't know unless someone tells her. This has two learning implications: Surgeons can't practice the skills needed to make holistic sense of the work on their own, and they must build new skills related to making sense of the work through others.

Benjamin Shestakofsky, now at the University of Pennsylvania, described a similar phenomenon at a pre-IPO start-up that used machine learning to match local laborers with jobs and that provided a platform for laborers and those hiring them to negotiate terms. At first the algorithms weren't making good matches, so managers in San Francisco hired people in the Philippines to manually create each match. And when laborers had difficulty with the platform—for instance, in using it to issue price quotes to those hiring, or to structure payments—the start-up managers outsourced the needed support to yet another distributed group of employees, in Las Vegas. Given their limited resources, the managers threw bodies at these problems to buy time while they sought the money and additional engineers needed to perfect the product. Delegation allowed the managers and engineers to focus on business development and writing code, but it deprived them of critical learning opportunities: It separated them from direct, regular input from customers—the laborers and the hiring contractors—about the problems they were experiencing and the features they wanted.

3. Learners are expected to master both old and new methods. Robotic surgery comprises a radically new set of techniques and technologies for accomplishing the same ends that traditional surgery seeks to achieve. Promising greater precision and ergonomics, it was simply added to the curriculum, and residents were expected to learn robotic as well as open approaches. But the curriculum didn't include enough time to

Idea in Brief

THE PROBLEM
The rush of intelligent machines and sophisticated analytics into many aspects of work means that trainees are losing opportunities to acquire skills through on-the-job learning (OJL).

THE OUTCOME
In medicine, policing, and other fields, people are finding rule-breaking ways to acquire needed expertise out of the limelight. This "shadow learning" is tolerated for the results it produces, but it can exact a personal and an organizational toll.

THE SOLUTION
In response, organizations should carefully uncover and study shadow learning; adapt practices that develop organizational, technological, and work designs that enhance OJL; and make intelligent machines part of the solution.

Close to a hundred years ago, the sociologist Robert Merton showed that when legitimate means are no longer effective for achieving a valued goal, deviance results.

learn both thoroughly, which often led to a worst-case outcome: The residents mastered neither. I call this problem *methodological overload*.

Shreeharsh Kelkar, at UC Berkeley, found that something similar happened to many professors who were using a new technology platform called edX to develop massive open online courses (MOOCs). EdX provided them with a suite of course-design tools and instructional advice based on fine-grained algorithmic analysis of students' interaction with the platform (clicks, posts, pauses in video replay, and so on). Those who wanted to develop and improve online courses had to learn a host of new skills—how to navigate the edX user interface, interpret analytics on learner behavior, compose and manage the course's project team, and more—while keeping "old school" skills sharp for teaching their traditional classes. Dealing with this tension was difficult for everyone, especially because the approaches were in constant flux: New tools, metrics, and expectations arrived almost daily, and instructors had to quickly assess and master them. The only people who handled both old and new methods well were those who were already technically sophisticated and had significant organizational resources.

4. Standard learning methods are presumed to be effective. Decades of research and tradition hold trainees in medicine to the *See one, do one, teach one* method, but as we've seen, it doesn't adapt well to robotic surgery. Nonetheless, pressure to rely on approved learning methods is so strong that deviation is rare: Surgical-training research, standard routines, policy, and senior surgeons all continue to emphasize traditional approaches to learning, even though the method clearly needs updating for robotic surgery.

Sarah Brayne, at the University of Texas, found a similar mismatch between learning methods and needs among police chiefs and officers in Los Angeles as they tried to apply traditional policing approaches to beat assignments generated by an algorithm. Although the efficacy of such "predictive policing" is unclear, and its ethics are controversial, dozens of police forces are becoming deeply reliant on it. The LAPD's PredPol system breaks the city up into 500-foot squares, or "boxes," assigns a crime probability to each one, and directs officers to those boxes accordingly. Brayne found that it wasn't always obvious to the officers—or to the police chiefs—when and how the former should follow their AI-driven assignments. In policing, the traditional and respected model for acquiring new techniques has been to combine a little formal instruction with lots of old-fashioned learning on the beat. Many chiefs therefore presumed that officers would mostly learn how to incorporate crime forecasts on the job. This dependence on traditional OJL contributed to confusion and resistance to the tool and its guidance. Chiefs didn't want to tell officers what to do once "in the box," because they wanted them to rely on their experiential knowledge and discretion. Nor did they want to irritate the officers by overtly reducing their autonomy and coming across as micromanagers. But by relying on the traditional OJL approach, they inadvertently sabotaged learning: Many officers never understood how to use PredPol or its potential benefits, so they wholly dismissed it—yet they were still held accountable for following its assignments. This wasted time, decreased trust, and led to miscommunication and faulty data entry—all of which undermined their policing.

Shadow Learning Responses

Faced with such barriers, shadow learners are bending or breaking the rules out of view to get the instruction and experience they need. We shouldn't be surprised. Close to a hundred years ago, the sociologist Robert Merton showed that when legitimate means are no longer effective for achieving a valued goal, deviance results. Expertise—perhaps the ultimate occupational goal—is no exception: Given the barriers I've described, we should expect people to find deviant ways to learn key skills. Their approaches are often ingenious and effective, but they can take a personal and an organizational toll: Shadow learners may be punished (for example, by losing practice opportunities and status) or cause waste and even harm. Still, people repeatedly take those risks, because their learning methods work well where approved means fail. It's almost always a bad idea to uncritically copy these deviant practices, but organizations do need to learn from them.

Following are the shadow learning practices that I and others have observed:

Seeking struggle. Recall that robotic surgical trainees often have

little time on task. Shadow learners get around this by looking for opportunities to operate near the edge of their capability and with limited supervision. They know they must struggle to learn, and that many attending physicians are unlikely to let them. The subset of residents I studied who did become expert found ways to get the time on the robots they needed. One strategy was to seek collaboration with attendings who weren't themselves seasoned experts. Residents in urology—the specialty having by far the most experience with robots—would rotate into departments whose attendings were less proficient in robotic surgery, allowing the residents to leverage the halo effect of their elite (if limited) training. The attendings were less able to detect quality deviations in their robotic surgical work and knew that the urology residents were being trained by true experts in the practice; thus they were more inclined to let the residents operate, and even to ask for their advice. But few would argue that this is an optimal learning approach.

What about those junior analysts who were cut out of complex valuations? The junior and senior members of one group engaged in shadow learning by disregarding the company's emerging standard practice and working together. Junior analysts continued to pull raw reports to produce the needed input, but they worked alongside senior partners on the analysis that followed.

In some ways this sounds like a risky business move. Indeed, it slowed down the process, and because it required the junior analysts to handle a wider range of valuation methods and calculations at a breakneck pace, it introduced mistakes that were difficult to catch. But the junior analysts developed a deeper knowledge of the multiple companies and other stakeholders involved in an M&A and of the relevant industry and learned how to manage the entire valuation process. Rather than function as a cog in a system they didn't understand, they engaged in work that positioned them to take on more-senior roles. Another benefit was the discovery that, far from being interchangeable, the software packages they'd been using to create inputs for analysis sometimes produced valuations of a given company that were billions of dollars apart. Had the analysts remained siloed, that might never have come to light.

Tapping frontline know-how. As discussed, robotic surgeons are isolated from the patient and so lack a holistic sense of the work, making it harder for residents to gain the skills they need. To understand the bigger picture, residents sometimes turn to scrub techs, who see the procedure in its totality: the patient's entire body; the position and movement of the robot's arms; the activities of the anesthesiologist, the nurse, and others around the patient; and all the instruments and supplies from start to finish. The best scrubs have paid careful attention during thousands of procedures. When residents shift from the console to the bedside, therefore, some bypass the attending and go straight to these "superscrubs" with technical questions, such as whether the intra-abdominal pressure is unusual, or when to clear the field of fluid or of smoke from cauterization. They do this despite norms and often unbeknownst to the attending.

And what about the start-up managers who were outsourcing jobs to workers in the Philippines and Las Vegas? They were expected to remain laser focused on raising capital and hiring engineers. But a few spent time with the frontline contract workers to learn how and why they made the matches they did. This led to insights that helped the company refine its processes for acquiring and cleaning data—an essential step in creating a stable platform. Similarly, some attentive managers spent time with the customer service reps in Las Vegas as they helped workers contend with the system. These "ride alongs" led the managers to divert some resources to improving the user interface, helping to sustain the start-up as it continued to acquire new users and recruit engineers who could build the robust machine learning systems it needed to succeed.

Redesigning roles. The new work methods we create to deploy intelligent machines are driving a variety of shadow learning tactics that restructure work or alter how performance is measured and rewarded. A surgical resident may decide early on that she isn't going to do robotic surgery as a senior physician and will therefore consciously minimize her robotic rotation. Some nurses I studied prefer the technical troubleshooting involved in robotic assignments, so they surreptitiously avoid open surgical work. Nurses who staff surgical procedures notice emerging preferences and skills and work around blanket staffing policies to accommodate them. People tacitly recognize and develop new roles that are better

aligned with the work—whether or not the organization formally does so.

Consider how some police chiefs reframed expectations for beat cops who were having trouble integrating predictive analytics into their work. Brayne found that many officers assigned to patrol PredPol's "boxes" appeared to be less productive on traditional measures such as number of arrests, citations, and FIs (field interview cards—records made by officers of their contacts with citizens, typically people who seem suspicious). FIs are particularly important in AI-assisted policing, because they provide crucial input data for predictive systems even when no arrests result. When cops went where the system directed them, they often made no arrests, wrote no tickets, and created no FIs.

Recognizing that these traditional measures discouraged beat cops from following PredPol's recommendations, a few chiefs sidestepped standard practice and publicly and privately praised officers not for making arrests and delivering citations but for learning to work with the algorithmic assignments. As one captain said, "Good, fine, but we are telling you where the probability of a crime is at, so sit there, and if you come in with a zero [no crimes], that is a success." These chiefs were taking a risk by encouraging what many saw as bad policing, but in doing so they were helping to move the law enforcement culture toward a future in which the police will increasingly collaborate with intelligent machines, whether or not PredPol remains in the tool kit.

Curating solutions. Trainees in robotic surgery occasionally took time away from their formal responsibilities to create, annotate, and share play-by-play recordings of expert procedures. In addition to providing a resource for themselves and others, making the recordings helped them learn, because they had to classify phases of the work, techniques, types of failures, and responses to surprises.

Faculty members who were struggling to build online courses while maintaining their old-school skills used similar techniques to master the new technology. EdX provided tools, templates, and training materials to make things easier for instructors, but that wasn't enough. Especially in the beginning, far-flung instructors in resource-strapped institutions took time to experiment with the platform, make notes and videos on their failures and successes, and share them informally with one another online. Establishing these connections was hard, especially when the instructors' institutions were ambivalent about putting content and pedagogy online in the first place.

Shadow learning of a different type occurred among the original users of edX—well-funded, well-supported professors at topflight institutions who had provided early input during the development of the platform. To get the support and resources they needed from edX, they surreptitiously shared techniques for pitching desired changes in the platform, securing funding and staff support, and so on.

Learning from shadow learners. Obviously shadow learning is not the ideal solution to the problems it addresses. No one should have to risk getting fired just to master a job. But these practices are hard-won, tested paths in a world where acquiring expertise is becoming more difficult and more important.

The four classes of behavior shadow learners exhibit—seeking struggle, tapping frontline know-how, redesigning roles, and curating solutions—suggest corresponding tactical responses. To take advantage of the lessons shadow learners offer, technologists, managers, experts, and workers themselves should:

- ensure that learners get opportunities to struggle near the edge of their capacity in real (not simulated) work so that they can make and recover from mistakes
- foster clear channels through which the best frontline workers can serve as instructors and coaches
- restructure roles and incentives to help learners master new ways of working with intelligent machines
- build searchable, annotated, crowdsourced "skill repositories" containing tools and expert guidance that learners can tap and contribute to as needed

The specific approach to these activities depends on organizational structure, culture, resources, technological options, existing skills, and, of course, the nature of the work itself. No single best practice will apply in all circumstances. But a large body of managerial literature explores each of these, and outside consulting is readily available.

More broadly, my research, and that of my colleagues, suggests three organizational strategies that may help leverage shadow learning's lessons:

Shadow learners find risky, rule-breaking ways to learn. Organizations that hope to compete in a world filling with increasingly intelligent machines should pay close attention to these "deviants."

1. Keep studying it. Shadow learning is evolving rapidly as intelligent technologies become more capable. New forms will emerge over time, offering new lessons. A cautious approach is critical. Shadow learners often realize that their practices are deviant and that they could be penalized for pursuing them. (Imagine if a surgical resident made it known that he sought out the least-skilled attendings to work with.) And middle managers often turn a blind eye to these practices because of the results they produce—as long as the shadow learning isn't openly acknowledged. Thus learners and their managers may be less than forthcoming when an observer, particularly a senior manager, declares that he wants to study how employees are breaking the rules to build skills. A good solution is to bring in a neutral third party who can ensure strict anonymity while comparing practices across diverse cases. My informants came to know and trust me, and they were aware that I was observing work in numerous work groups and facilities, so they felt confident that their identities would be protected. That proved essential in getting them to open up.

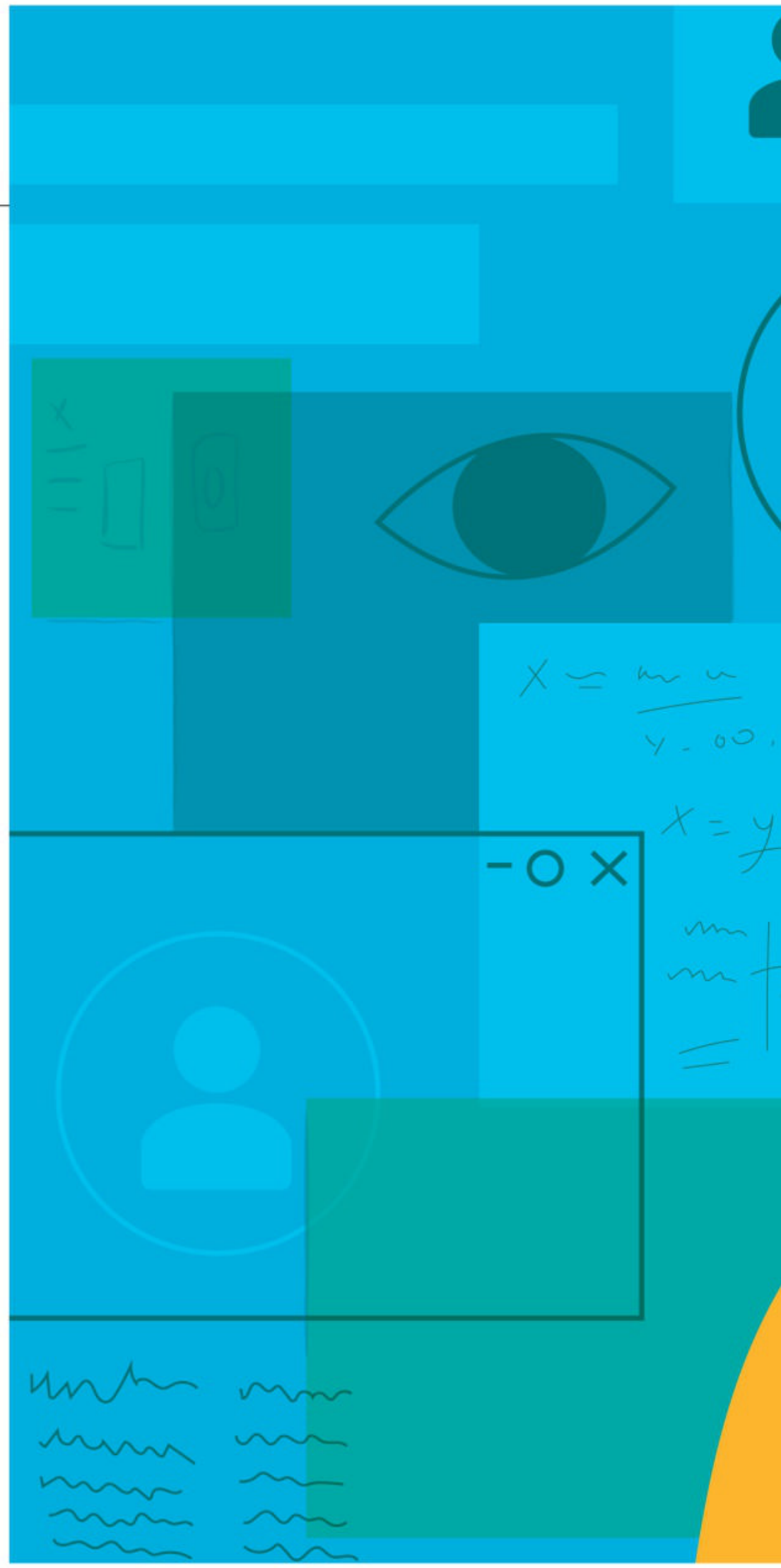
2. Adapt the shadow learning practices you find to design organizations, work, and technology. Organizations have often handled intelligent machines in ways that make it easier for a single expert to take more control of the work, reducing dependence on trainees' help. Robotic surgical systems allow senior surgeons to operate with less assistance, so they do. Investment banking systems allow senior partners to exclude junior analysts from complex valuations, so they do. All stakeholders should insist on organizational, technological, and work designs that improve productivity and enhance on-the-job learning. In the LAPD, for example, this would mean moving beyond changing incentives for beat cops to efforts such as redesigning the PredPol user interface, creating new roles to bridge police officers and software engineers, and establishing a cop-curated repository for annotated best practice use cases.

3. Make intelligent machines part of the solution. AI can be built to coach learners as they struggle, coach experts on their mentorship, and connect those two groups in smart ways. For example, when Juho Kim was a doctoral student at MIT, he built ToolScape and LectureScape, which allow for crowdsourced annotation of instructional videos and provide clarification and opportunities for practice where many prior users have paused to look for them. He called this *learnersourcing*. On the hardware side, augmented reality systems are beginning to bring expert instruction and annotation right into the flow of work. Existing applications use tablets or smart glasses to overlay instructions on work in real time. More-sophisticated intelligent systems are expected soon. Such systems might, for example, superimpose a recording of the best welder in the factory on an apprentice welder's visual field to show how the job is done, record the apprentice's attempt to match it, and connect the apprentice to the welder as needed. The growing community of engineers in these domains have mostly been focused on formal training, and the deeper crisis is in on-the-job learning. We need to redirect our efforts there.

FOR THOUSANDS OF YEARS, advances in technology have driven the redesign of work processes, and apprentices have learned necessary new skills from mentors. But as we've seen, intelligent machines now motivate us to peel apprentices away from masters, and masters from the work itself, all in the name of productivity. Organizations often unwittingly choose productivity over considered human involvement, and learning on the job is getting harder as a result. Shadow learners are nevertheless finding risky, rule-breaking ways to learn. Organizations that hope to compete in a world filling with increasingly intelligent machines should pay close attention to these "deviants." Their actions provide insight into how the best work will be done in the future, when experts, apprentices, and intelligent machines work, and learn, together.

HBR Reprint R1905K

Matt Beane *is an assistant professor of technology management at the University of California, Santa Barbara, and a digital fellow with the Stanford Digital Economy Lab and MIT's Initiative on the Digital Economy.*



ORIGINALLY PUBLISHED JANUARY–FEBRUARY 2016

Algorithms Need Managers, Too

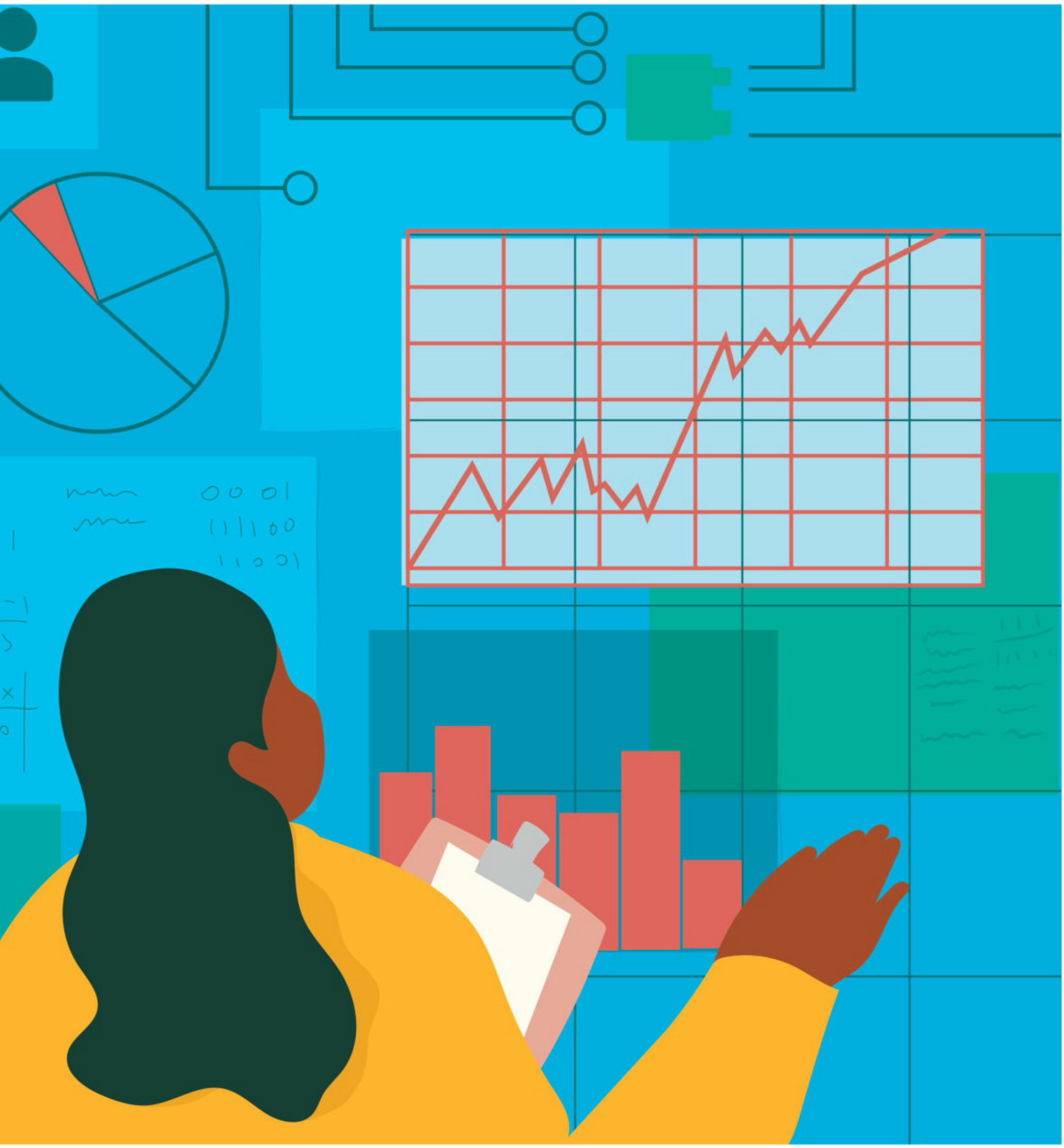
Know how to get the most out of your predictive tools.

→ by MICHAEL LUCA, JON KLEINBERG, and SENDHIL MULLAINATHAN

MOST MANAGERS' JOBS involve making predictions. When HR specialists decide whom to hire, they're predicting who will be most effective. When marketers choose which distribution channels to use, they're predicting where a product will sell best. When VCs determine whether to fund a start-up, they're predicting whether it will succeed. To make these and myriad other business predictions,companies today are turning more and more to computer algorithms, which perform step-by-step analytical operations at incredible speed and scale.

Algorithms make predictions more accurate—but they also create risks of their own, especially if we do not understand them. High-profile examples abound. When Netflix

Illustration by MICHELLE D'URBANO



ran a million-dollar competition to develop an algorithm that could identify which movies a given user would like, teams of data scientists joined forces and produced a winner. But it was one that applied to DVDs—and as Netflix's viewers transitioned to streaming movies, their preferences shifted in ways that didn't match the algorithm's predictions.

Another example comes from social media. Today many sites deploy algorithms to decide which ads and links to show users. When these algorithms focus too narrowly on maximizing user click-throughs, sites become choked with low-quality "click-bait" articles. Click-through rates rise, but overall customer satisfaction may plummet.

Problems like these aren't inevitable. In our work designing and implementing algorithms and identifying new data sources with a range of organizations, we have seen that the source of difficulty often isn't bugs in the algorithms; it's bugs in the way we interact with them. To avoid missteps, managers need to understand what algorithms do well—what questions they answer and what questions they do not.

Why Do Smart Algorithms Lead Us Astray?

As a growing body of evidence shows, humanizing algorithms makes us more comfortable with them. This can be useful if, for example, you're designing an automated call function. A real person's voice is more likely than an electronic voice to get people to listen. The fundamental problem, however, is that people treat algorithms and the machines that run them the same way they'd treat an employee, supervisor, or colleague. But algorithms behave very differently from humans, in two important ways:

Algorithms are extremely literal. In the latest Avengers movie, Tony Stark (also known as Iron Man) creates Ultron, an artificial-intelligence defense system tasked with protecting Earth. But Ultron interprets the task literally, concluding that the best way to save Earth is to destroy all humans. In many ways, Ultron behaves like a typical algorithm: It does exactly what it's told—and ignores every other consideration. We get into trouble when we don't manage algorithms carefully.

The social media sites that were suddenly swamped with click-bait fell into a similar trap. Their overall goal was clear: Provide content that would be most appealing and engaging to users. In communicating it to the algorithm, they came up with a set of instructions that seemed like a good proxy—find items that users will click on the most. And it's not a bad proxy: People typically click on content because it interests them. But making selections solely on the basis of clicks quickly filled sites with superficial and offensive material that hurt their reputation. A human would understand that the sites' designers meant "Maximize quality as measured by clicks," not "Maximize clicks even at the expense of quality." An algorithm, on the other hand, understands only what it is explicitly told.

Algorithms are black boxes. In Shakespeare's *Julius Caesar,* a soothsayer warns Caesar to "beware the ides of March." The recommendation was perfectly clear: Caesar had better watch out. Yet at the same time it was completely incomprehensible. Watch out for what? Why? Caesar, frustrated with the mysterious message, dismissed the soothsayer, declaring, "He is a dreamer; let us leave him." Indeed, the ides of March turned out to be a bad day for the ruler. The problem was that the soothsayer provided *incomplete* information. And there was no clue to what was missing or how important that information was.

Like Shakespeare's soothsayer, algorithms often can predict the future with great accuracy but tell you neither what will cause an event nor why. An algorithm can read through every *New York Times* article and tell you which is most likely to be shared on Twitter without necessarily explaining why people will be moved to tweet about it. An algorithm can tell you which employees are most likely to succeed without identifying which attributes are most important for success.

Recognizing these two limitations of algorithms is the first step to managing them better. Now let's look at other steps you can take to leverage them more successfully.

Be Explicit About All Your Goals

Everyone has objectives and directives, but we also know that the end doesn't always justify the means. We understand that there are soft (often unspoken) goals and trade-offs. We may turn down a little profit today for a gain in reputation tomorrow. We may strive for equality—even if it causes organi-

While people understand soft goals and trade-offs, algorithms will pursue a specified objective single-mindedly.

zational pain in the short term. Algorithms, on the other hand, will pursue a specified objective single-mindedly. The best way to mitigate this is to be crystal clear about everything you want to achieve.

If you care about a soft goal, you need to state it, define it, and quantify how much it matters. To the extent that soft goals are difficult to measure, keep them top of mind when acting on the results from an algorithm.

At Google (which has funded some of our research on other topics), a soft-goal problem emerged with an algorithm that determines which ads to display. Harvard professor Latanya Sweeney unearthed it in a study. She found that when you typed names that were typically African American, like "Latanya Farrell," into Google, you were shown ads offering to investigate possible arrest records, but not when you searched on names like "Kristen Haring." Google's hard goal of maximizing clicks on ads had led to a situation in which its algorithms, refined through feedback over time, were in effect defaming people with certain kinds of names. It happened because people who searched for particular names were more likely to click on arrest records, which led these records to appear even more often, creating a self-reinforcing loop. This probably was not the intended outcome, but without a soft goal in place, there was no mechanism to steer the algorithm away from it.

We recently saw the importance of soft goals in action. One of us was working with a West Coast city to improve the efficiency of its restaurant inspections. For decades, the city had been doing them mostly at random but giving more-frequent scrutiny to places with prior violations. Choosing which establishments to inspect is an ideal job for an algorithm, however. Our algorithm found many more variables—not just past violations—to be predictive. The result was that the health department could identify probable offenders more easily and then find actual violations with far fewer inspections.

The officials loved the idea of making the process more efficient and wanted to move toward implementation. We asked if there were any questions or concerns. After an awkward silence, one person raised her hand. "I don't know how to bring this up," she said. "But there's an issue we should discuss." She explained that in some neighborhoods with tighter quarters, there tended to be more violations. These neighborhoods also happened to be home to higher percentages of minority residents with lower incomes. She did not want these neighborhoods to be excessively targeted by the algorithm. She was expressing a soft goal related to fairness. Our simple solution was to incorporate that objective into the algorithm by setting a ceiling on the number of inspections within each area. This would achieve the hard goal, identifying the restaurants most likely to have problems, while still respecting the soft one, ensuring that poor neighborhoods were not singled out.

Notice the extra step that allowed us to bake in soft goals: giving everyone an opportunity to articulate any concerns. We find that people often formulate soft goals as concerns, so asking for them explicitly facilitates more open and fruitful discussion. It's also critical to give people license to be candid and up-front—to say things that they wouldn't normally. This approach can surface a variety of issues, but the ones we see most commonly relate to fairness and to the handling of sensitive situations.

With a core objective and a list of concerns in hand, the designer of the algorithm can then build trade-offs into it. Often that may mean extending the

Idea in Brief

THE PROBLEM
Algorithms are essential tools for planning, but they can easily lead decision makers astray.

THE CAUSES
All algorithms share two characteristics: They're literal, meaning that they'll do exactly what you ask them to do. And they're black boxes, meaning that they don't explain why they offer particular recommendations.

THE SOLUTION
When formulating algorithms, be explicit about all your goals. Consider long-term implications of the data you examine. And make sure you choose the right data inputs.

Algorithms focus on the data at hand—and that data often pertains just to short-term outcomes.

objective to include multiple outcomes, weighted by importance.

Minimize Myopia

A popular consumer packaged goods company was purchasing products cheaply in China and selling them in the United States. It selected these products after running an algorithm that forecast which ones would sell the most. Sure enough, sales took off and cruised along nicely—until several months later, when customers started to return the items.

As it happens, the surprisingly high and steady return rate could have been predicted (even though the algorithm had failed to foresee it). The company obviously cared about quality, but it hadn't translated that interest into an algorithm that carefully projected consumer satisfaction; instead it had asked the algorithm to focus narrowly on sales. Ultimately, the company's new approach was to become great at forecasting not just how well products would sell but also how much people would enjoy and keep their products. The firm now looks for offerings that customers will rave about on Amazon and other platforms, and the product return rate has plummeted.

This company ran into a common pitfall of dealing with algorithms: Algorithms tend to be myopic. They focus on the data at hand—and that data often pertains to short-term outcomes. There can be a tension between short-term success and long-term profits and broader corporate goals. Humans implicitly understand this; algorithms don't unless you tell them to.

This problem can be solved at the objective-setting phase by identifying and specifying long-term goals. But when acting on an algorithm's predictions, managers should also adjust for the extent to which the algorithm is consistent with long-term aims.

Myopia is also the underlying weakness of programs that produce low-quality content by seeking to maximize click-throughs. The algorithms are optimizing for a goal that can be measured in the moment—whether a user clicks on a link—without regard to the longer-range and more important goal of keeping users satisfied with their experience on the site.

Nearsightedness can similarly be an issue with marketing campaigns. Consider a run-of-the-mill Gap advertising campaign with Google. It would most likely lead to a spike in visits to Gap.com—because Google's algorithm is good at predicting who will click on an ad. The issue is, the real goal is increasing sales—not increasing website visits. To address this, advertising platforms can collect sales data through a variety of channels, such as partnerships with payment systems, and incorporate it into their algorithms.

What's more, website visits are a short-term behavior, whereas the long-term impact of advertisements includes the downstream effects on brand image and repeat business. While perfect data on such effects is hard to find, careful data audits can help a lot. Managers should systematically list all internal and external data that may be relevant to the project at hand. With a Google campaign, the Gap's marketers could begin by laying out all their objectives—high sales, low returns, good reputation, and so on—and then spell out ways to measure each. Product returns, online reviews, and searches for the term "Gap" would all be great metrics. The best algorithm could then build predictions from a combination of all these features, calibrating for their relative importance.

Choose the Right Data Inputs

Let's return to the example of health departments that are trying to identify restaurants at risk for causing food-borne illness. As mentioned earlier, cities historically have inspected either randomly or on the basis of prior inspection results. Working with Yelp, one of us helped the city of Boston use online reviews to determine which restaurants were most likely to violate local health codes, creating an algorithm that compared the text in reviews with historical inspection data. By applying it, the city identified the same number of violations as usual, but with 40% fewer inspectors—a dramatic increase in efficiency.

This approach worked well not just because we had a lot of restaurants to look at but because Yelp reviews provided a great set of data—something cities hadn't given much thought to. A Yelp review contains many words and a variety of information. The data is also diverse, because it's drawn from different sources. In short, it's quite unlike the inspector-created data cities were accustomed to working with.

When choosing the right data resources, keep in mind the following:

Wider is better. One trap companies often fall into is thinking of big data as simply a lot of records—for example, looking at one million customers instead of 10,000. But this is only half the picture. Imagine your data organized into a table, with a row for each customer. The number of customers is the length of the table. The amount you know about each customer determines the width—how many features are recorded in each row. And while increasing the length of the data will improve your predictions, the full power of big data comes from gathering wide data. Leveraging comprehensive information is at the heart of prediction. Every additional detail you learn about an outcome is like one more clue, and it can be combined with clues you've already collected. Text documents are a great source of wide data, for instance; each word is a clue.

Diversity matters. A corollary to this is that data should be diverse, in the sense that the different data sources should be relatively unrelated to one another. This is where extra predictive power comes from. Treat each data set like a recommendation from a friend. If the data sets are too similar, there won't be much marginal gain from each additional one. But if each data set has a unique perspective, a lot more value is created.

Understand the Limitations

Knowing what your algorithm can't tell you is just as important as knowing what it can. It's easy to succumb to the misguided belief that predictions made in one context will apply equally well in another. That's what prevented the 2009 Netflix competition from yielding more benefit to the company: The algorithm that accurately forecast which DVD a person would want to order in the mail wasn't nearly as good at pinpointing which movie a person would want to stream right now. Netflix got useful insights and good publicity from the contest, but the data it collected on DVDs did not apply to streaming.

Algorithms use existing data to make predictions about what might happen with a slightly different setting, population, time, or question. In essence, you're transferring an insight from one context to another. It's a wise practice, therefore, to list the reasons why the algorithm might not be transferable to a new problem and assess their significance. For instance, a health-code violation algorithm based on reviews and violations in Boston may be less effective in Orlando, which has hotter weather and therefore faces different food safety issues.

Also remember that correlation still doesn't mean causation. Suppose that an algorithm predicts that short tweets will get retweeted more often than longer ones. This does not in any way suggest that you should shorten your tweets. This is a prediction, not advice. It works as a prediction because there are many other factors that correlate with short tweets that make them effective. This is also why it fails as advice: Shortening your tweets will not necessarily change those other factors.

Consider the experiences of eBay, which had been advertising through Google for years. EBay saw that people who viewed those ads were more likely to shop at it than people who did not. What it didn't see was whether the advertisements (which were shown millions of times) were causing people to come to its site. After all, the ads were deliberately shown to likely eBay shoppers. To separate correlation from causation, eBay ran a large experiment in which it randomly advertised to some people and not others. The result? It turns out that the advertisements were for the most part useless, because the people who saw them already knew about eBay and would have shopped there anyway.

ALGORITHMS CAPABLE OF making predictions do not eliminate the need for care when drawing connections between cause and effect; they are not a replacement for controlled experiments. But what they can do is extremely powerful: identifying patterns too subtle to be detected by human observation, and using those patterns to generate accurate insights and inform better decision making. The challenge for us is to understand their risks and limitations and, through effective management, unlock their remarkable potential.

HBR Reprint R1601H

Michael Luca *is the Lee J. Styslinger III Associate Professor of Business Administration at Harvard Business School and author of* The Power of Experiments: Decision Making in a Data-Driven World *(MIT Press, 2020).* ***Jon Kleinberg*** *is a professor of computer science at Cornell University and the coauthor of the textbooks* Algorithm Design *(with Éva Tardos) and* Networks, Crowds, and Markets *(with David Easley).* ***Sendhil Mullainathan*** *is a professor of behavioral and computational sciences at the University of Chicago and the coauthor (with Eldar Shafir) of* Scarcity: Why Having Too Little Means So Much.

Quick Takes

ANDRIY ONUFRIYENKO/GETTY IMAGES

1. The Future of AI Will Be About Less Data, Not More

→ by H. JAMES WILSON, PAUL R. DAUGHERTY, and CHASE DAVENPORT

COMPANIES CONSIDERING how to invest in AI capabilities should first understand that over the next five years applications and machines will become less artificial and more intelligent. They will rely less on bottom-up big data and more on top-down reasoning that more closely resembles the way humans approach problems and tasks. This general reasoning ability will enable AI to be more broadly applied than ever, creating opportunities for early adopters even in businesses and activities to which it previously seemed unsuited.

In the recent past, AI advanced through deep learning and machine learning, building up systems from the bottom by training them on mountains of data. For instance, driverless vehicles are trained on as many traffic situations as possible. But these data-hungry neural networks, as they are called, have serious limitations. They especially have trouble handling "edge" cases—situations where little data exists. A driverless car that can handle crosswalks, pedestrians, and traffic has trouble processing anomalies like children dressed in unusual Halloween costumes, weaving across the street after dusk.

Many systems are also easily stumped. The iPhone X's facial recognition system doesn't recognize "morning faces"—a user's puffy, haggard look on first awakening. Neural networks have beaten chess champions and triumphed at the ancient Japanese game of Go, but turn an image upside down or slightly alter it and the network may misidentify it. Or it may provide "high confidence" identifications of unrecognizable objects.

Data-hungry systems also face business and ethical constraints. Not every company has the volume of data necessary to build unique capabilities using neural networks. Using huge amounts of citizens' data also raises privacy issues that may lead to more government action like the European Union's General Data Protection Regulation (GDPR), which imposes stringent requirements on the use of individuals' personal data. Further, these systems are black boxes—it's not clear how they use input data to arrive at outputs like actions or decisions. This leaves them open to manipulation by bad actors (like the Russians in

In the future, we will have top-down systems that, like humans, are more innately intelligent.

the 2016 U.S. presidential election), and when something goes embarrassingly wrong, organizations are hard put to explain why.

In the future, however, we will have top-down systems that don't require as much data and are faster, more flexible, and, like humans, more innately intelligent. Several companies and organizations are already putting these more natural systems to work. To craft a vision of where AI is heading in the next several years, and then plan investments and tests accordingly, companies should look for developments in four areas:

More-efficient robot reasoning. When robots have a conceptual understanding of the world, as humans do, it is easier to teach them things, using far less data. Vicarious, a Union City, California, start-up whose investors include Mark Zuckerberg, Jeff Bezos, and Marc Benioff, is working to develop artificial general intelligence (AGI) for robots, enabling them to generalize from few examples.

Consider those jumbles of letters and numerals that websites use to determine whether you're a human or a robot. Called CAPTCHAs (Completely Automated Public Turing tests to tell Computers and Humans Apart), they are easy for humans to solve and hard for computers. Drawing on computational neuroscience, researchers at Vicarious have developed a model that can break through CAPTCHAs at a far higher rate than deep neural networks and with 300-fold more data efficiency. To parse CAPTCHAs with almost 67% accuracy, the Vicarious model required only five training examples per character, while a state-of-the-art deep neural network required a 50,000-fold larger training set of actual CAPTCHA strings. Such models, with their ability to train faster and generalize more broadly than AI approaches commonly used today, are putting us on a path toward robots that have a humanlike conceptual understanding of the world.

Ready expertise. Modeling what a human expert would do in the face of high uncertainty and little data, top-down AI can beat data-hungry approaches for designing and controlling many varieties of factory equipment. Siemens is using top-down AI to control the highly complex combustion process in gas turbines, where air and gas flow into a chamber, ignite, and burn at temperatures as high as 1,600 degrees Celsius. The volume of emissions created and ultimately how long the turbine will continue to operate depends on the interplay of numerous factors, from the quality of the gas to airflow and internal and external temperature.

Using bottom-up machine-learning methods, the gas turbine would have to run for a century before producing enough data to begin training. Instead, Siemens researchers Volar Sterzing and Steffen Udluft used methods that required little data in the learning phase for the machines. The monitoring system that resulted makes fine adjustments that optimize how the turbines run in terms of emissions and wear, continuously seeking the best solution in real time, much like an expert knowledgeably twirling multiple knobs in concert.

Common sense. A variety of organizations are working to teach machines to navigate the world using common sense—to understand everyday objects and actions, communicate naturally, handle unforeseen situations, and learn from experiences. But what comes naturally to humans, without explicit training or data, is fiendishly difficult for machines. Says Oren Etzioni, the CEO of the Allen Institute for Artificial Intelligence (AI2), "No AI system currently deployed can reliably answer a broad range of simple questions, such as, 'If I put my socks in a drawer, will they still be in there tomorrow?' or 'How can you tell if a milk carton is full?'"

To help define what it means for machines to have common sense, AI2 is developing a portfolio of tasks against which progress can be measured. The Defense Advanced Research Projects Agency (DARPA) is investing $2 billion in AI research. In its Machine Common Sense program, researchers will create models that mimic core domains of human cognition, including "the domains of *objects* (intuitive physics), *places* (spatial navigation), and *agents* (intentional actors)." Researchers at Microsoft and McGill University have jointly developed a system that has shown great promise for untangling ambiguities in natural language, a problem that requires diverse forms of inference and knowledge.

Better bets. Humans routinely, and often effortlessly, sort through probabilities and act on the likeliest, even with relatively little prior

experience. Machines are now being taught to mimic such reasoning through the application of Gaussian processes—probabilistic models that can deal with extensive uncertainty, act on sparse data, and learn from experience. Alphabet, Google's parent company, launched Project Loon, designed to provide internet service to underserved regions of the world through a system of giant balloons hovering in the stratosphere. Their navigational systems employ Gaussian processes to predict where in the stratified and highly variable winds aloft the balloons need to go. Each balloon then moves into a layer of wind blowing in the right direction, arranging themselves to form one large communication network. The balloons can not only make reasonably accurate predictions by analyzing past flight data but also analyze data during a flight and adjust their predictions accordingly.

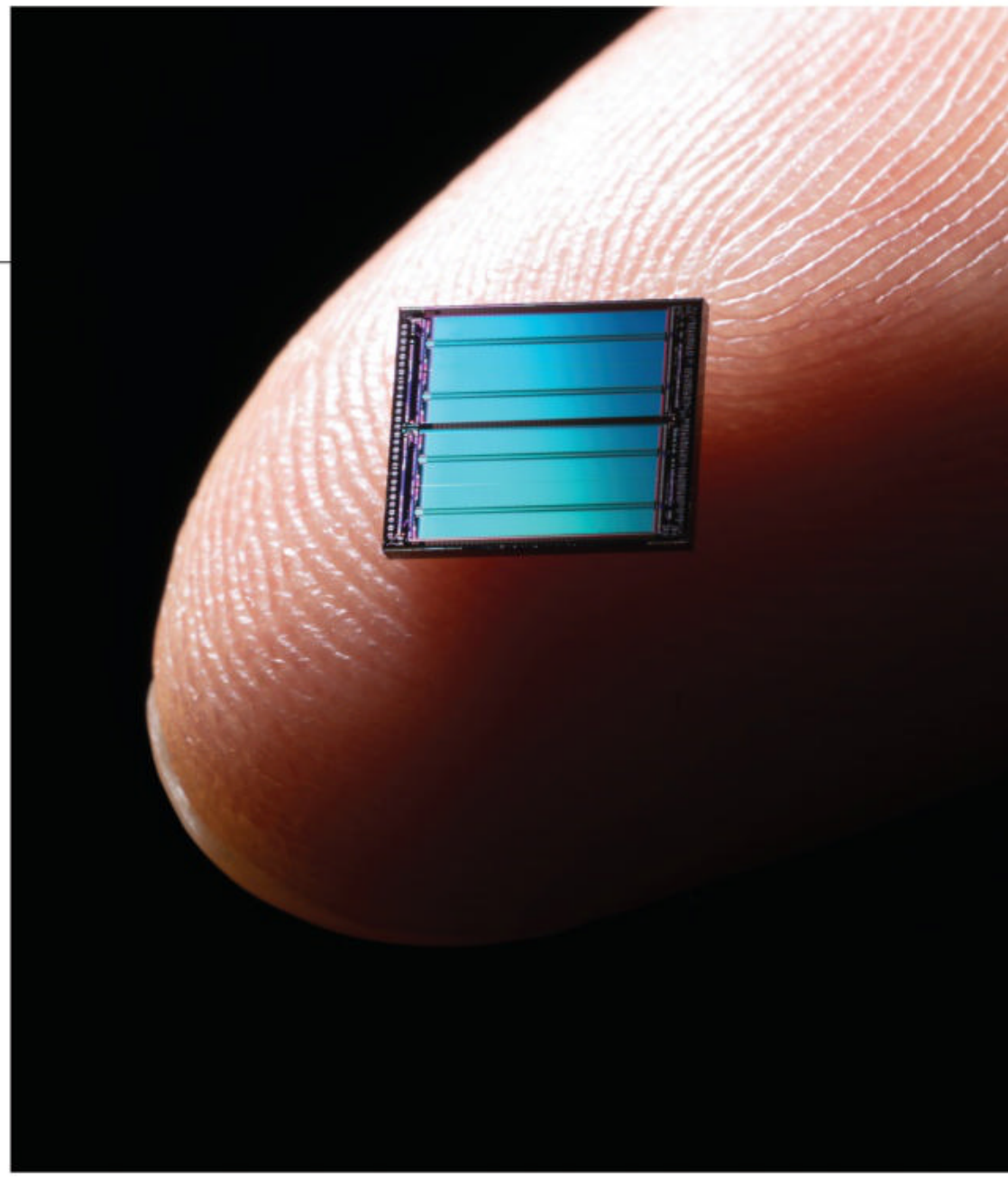
Such Gaussian processes hold great promise. They don't require massive amounts of data to recognize patterns, the computations required for inference and learning are relatively easy, and if something goes wrong, its cause can be traced, unlike the black boxes of neural networks.

Although all these advances are relatively recent, they hark back to the very beginnings of AI in the 1950s, when several researchers began to pursue top-down models for mimicking human intelligence. But when progress proved elusive and the rich potential for bottom-up machine-learning methods became apparent, the top-down approach was largely abandoned. Today, however, through powerful new research and computational techniques, top-down AI has been reborn. As its great promise begins to be fulfilled, smart companies will put their money where the mind is.

HBR Reprint H04QN6

Originally published on HBR.org January 14, 2019

H. James Wilson *is the global managing director of information technology and business research at Accenture Research. Follow him on Twitter @hjameswilson. He is a coauthor, with Paul R. Daugherty, of* Human + Machine: Reimagining Work in the Age of AI *(Harvard Business Review Press, 2018).* ***Paul R. Daugherty*** *is Accenture's group chief executive of technology and chief technology officer. He is a coauthor, with H. James Wilson, of* Human + Machine: Reimagining Work in the Age of AI. *Follow him on Twitter @pauldaugh.* ***Chase Davenport*** *is a San Francisco–based technology researcher.*



MIRAGEC/GETTY IMAGES

2. The Secret to AI Is People

→ by NADA R. SANDERS and JOHN D. WOOD

TOO MANY BUSINESS leaders still believe that AI is just another plug-and-play incremental technological investment. In reality, gaining a competitive advantage through AI requires organizational transformation of the kind exemplified by companies leading in this era: Google, Haier, Apple, Zappos, and Siemens. These companies don't just have better technology—they have transformed the way they do business so that human resources can be augmented with machine powers.

How do they do it? To find out, we conducted a multistage study over five years, beginning with a survey of senior managers and execu-

Competing in the age of AI is about having new organizational structures that use technology to bring out the best in people.

tives, followed by interviews and surveys across a wide range of industries to identify technology implementation strategies and barriers, and in-depth studies of five leading organizations. Our key takeaway is counterintuitive. Competing in the age of AI is not about being technology driven per se—it's about having new organizational structures that use technology to bring out the best in people. The secret to making this work, we learned, is the business model itself, where machines and humans are integrated to complement each other. Machines do repetitive and automated tasks and will always be more precise and faster. However, those uniquely human skills of creativity, care, intuition, adaptability, and innovation are increasingly imperative to success. These human skills cannot be "bot-sourced," a term we use to characterize when a business process traditionally carried out by humans is delegated to an automated process like a robot or an algorithm.

How Do Leaders Get the Most out of AI?

From our research we have developed a four-layer framework, the Four 1 model, that shows organizational leaders how they can create a human-centric organization with superhuman intelligence. The four layers—intentionality, integration, implementation, and indication—are not "steps," which would imply a sequential progression. They must be stacked all together or the use of AI will fail to deliver a sustainable competitive advantage. Here's how it works.

Intentionality. The first layer of the Four I model is *intentionality* of purpose, beyond the mere pursuit of profits. An intentional organization knows why it matters to the world, not just to its shareholders. A good example of intentionality in the use of AI comes from Siemens, which evolved from a shareholder-profit-maximizing power generation and transmission company into a leading provider of electrification, automation, and digitalization solutions with energy-efficient, resource-saving technologies driven by AI and the Internet of Things (IoT) in service to society. This cultural shift toward a higher human-centric purpose impacted not just marketing and product design but also the strategic decision to, as Scott D. Anthony, Alasdair Trotter, and Evan I. Schwartz wrote for HBR, "divest its core oil and gas business and redeploy the capital to its Digital Industries unit and Smart Infrastructure business focused on energy efficiency, renewable power storage, distributed power, and electric vehicle mobility." While financial performance and shareholder value will always be important, creating human-centered, technology-powered organizations will actually drive financial performance in the age of AI.

To that end, Siemens is launching a combination of hardware and software that enables AI throughout its Totally Integrated Automation (TIA) architecture, an approach that aligns Siemens's mission with its AI strategy. The TIA architecture uses AI as a bridge that spans from corporate headquarters out to industrial end users. Siemens's proprietary MindSphere is a cloud-based IoT operating platform that reaches into Siemens's industrial user-operated controller and field device products. The MindSphere's neural processing unit module allows human users to benefit from Siemens's in-house AI capabilities, while also enabling human users to impart their own experience to train the machines. According to Siemens's Factory Automation Specialist Colm Gavin, "With artificial intelligence we are able to train, recognize, and adjust to allow more flexible machinery, because, do we want 10 machines to package 10 different types of products, or a tool that accommodates different packages and different sizes and automatically adjusts to the new format?" Smarter machinery with TIA architecture leverages AI to advance the company's intentionality, while increasing flexibility, quality, efficiency, and cost-effectiveness for its end users.

Creating human-centered, technology-powered organizations will actually drive financial performance in the age of AI.

Alternatively, a negative example of the relationship between intentionality and AI is illustrated by recent issues confronting Facebook. Facebook's mission, "to give people the power to build community and bring the world closer together," sounds noble. Yet recent use of its AI has raised concerns from advertisers and civil rights groups alike. The social media giant has struggled to align its mission with its use of AI, which seems to have the opposite effect: Facebook's content "feed" is driven by algorithms that prioritize inflammatory, misleading, and socially divisive content. Facebook's use of AI seems to drive social division, which is antithetical to its purpose as a social media company, and is having financial consequences. Because its algorithms have promoted disinformation, violence, and incendiary content, major advertisers are now cutting ties with Facebook, dealing a strong blow to the company that derives 98% of its income from ad revenue. Some of the largest brands in the world, including Coca-Cola and Unilever, pulled advertisements from Facebook for promoting content antithetical to their brand's values, resulting in a one-day drop of 8.3% in market value, or $56 billion.

Integration. The second layer of the model is *integration* of human and AI resources across the organization. To lead in the technology era, companies must shift away from silos to organizational structures with flexible teams that integrate people horizontally and vertically, from product creation to strategic decision-making. As one executive we spoke with explained, before the AI shift, workers needed to have deep knowledge of a narrow area. Today, deep analytical content can come from AI. Workers need to be able to synthesize information, which means collaborating across functions and working in cross-functional teams. To foster innovation and adaptability, organizations need to transition from rigid hierarchies to flexible, agile, and flatter structures. Google, Haier, and Zappos may have differences in their organizational structures, but the common elements are flatness and fluidity. The recommended structure is more like a playground for smart, talented people to generate customer-centric products. Employees have fluid roles in cross-functional teams around problems rather than individual roles and responsibilities. These teams spontaneously form when problems arise, then dissolve when the work is done, reallocating human resources as needed.

The other side of this—which can easily be forgotten—is that human and AI teams should also be structured in an integrated manner. This allows humans to transcend their ordinary cognitive limitations without placing unreasonable reliance on a robot to perform human tasks that require high degrees of care and skill. An example comes from the medical context, where AI offers tremendous potential not as a substitute for, but as a supplement to, physician-driven care. Recent research in the journal *Nature* found that "good quality AI-based support of clinical decision-making improves diagnostic accuracy over that of either AI or physicians alone." This means high-stakes, highly skilled human decision-making can benefit from AI so long as it is integrated properly within the human decision-making context.

Implementation. The third layer of the model is *implementation*. Implementation requires engaging human talent, tolerating risk, and incentivizing cross-functional coordination. An executive at a large pharmaceutical we spoke with said, "You have to get people to believe in the technology." We saw this in another company we spoke with when we learned that despite having integrated AI, managers were modifying the output values from the algorithm to fit their own expectations. Others in the same company would simply follow the old decision-making routine, altogether ignoring the data provided by algorithms. Therefore, human behavior is central to implementing AI.

Top-performing companies spent significant time communicating with employees and educating them so that the human talent understood how machines made their jobs easier, not obsolete. To build trust in AI, it is imperative for leaders to communicate their vision transparently, explaining the goal, the changes needed, how it will be rolled out, and over what timeline. Beyond communication, leaders can inoculate their workforce against fear of AI by arranging for visits to other companies that have undergone similar transformations, providing a model for workers to see with their own eyes how the technology is used.

We saw many approaches to this in our research. Pilot projects where technology is rolled out in a limited scope give workers some ownership over the adoption process. Giving workers an opportunity to tinker with the technology before a final adoption decision is made eases the transition. Financial services firm Capital One even created an internal training institute called Capital One University that offers professional training programs to promote a broader understanding of analytics throughout the organization's culture.

Indication. The fourth layer of the model is *indication*, or performance measurement. Ultimately, success and progress need to be measured, and leading companies have moved from traditional productivity measures to aspirational metrics. Using the right indicators can drive improvements and help a business focus on what they deem important. Aspirational metrics that incentivize innovation and creativity encourage employees to exercise those uniquely human traits. The lesson is to be careful what you measure. Monitoring the wrong performance indicator has a strong tendency to lead to the proverbial tail wagging the dog. Humans are clever, and if incentives are not properly aligned with intelligently designed performance metrics, human workers will resort to lazy, clever, and cynical hacks to game the system, maximizing the appearance of performance under one measure while actually failing to deliver the output that management was actually hoping for when they implemented that measure.

Most companies use key performance indicators (KPIs,) but in our research we saw that successful companies more often used objectives and key results (OKRs). What we learned was that KPIs by themselves don't encompass strategic and ambitious goals needed in the age of AI and they don't motivate to reach for the sky. The goal of OKRs is to precisely define how to achieve ambitious objectives where failure is imminently possible, through concrete, measurable specifications. They encourage creative, novel, and aspirational performance by showing progress toward a goal even if the goal itself is unattained. Google famously started using OKRs in 1999, a change some even credit as a critical element of Google's success. At Google, OKRs have helped develop transparency. Everybody knows the company's goals, what everyone is doing, how they have done in the past, the trajectory they are on, and how they are getting to where they want to go.

Building Companies on Superhuman Intelligence

Our research shows that AI is so much more than just the latest incremental improvement in existing technology. However, deploying it effectively takes leadership and coordination across all sectors of a company. Unlocking the full potential of an organization's human resources by adopting AI strategically requires revisiting the very structure of the company and how it measures its progress toward fulfilling its mission. These issues are essential to the identity of a company, and modifications here are fraught with insecurity and risk, but this is a risk needed to compete in the age of AI. Intentionality, integration, implementation, and indication must be layered to create a human-centric enterprise governed by superhuman intelligence. Achieving this requires talent at all levels to have systems thinking and to understand how the work being done meshes with that of others elsewhere in the organization, how it meets customer needs, and how it impacts the company's strategy and financial picture. By following the Four I model, companies can unlock superhuman intelligence without losing the human touch.

We were surprised to discover how few organizations have unlocked this secret. But we were encouraged by the progress of the ones that had. With this model, we hope, more companies can create the conditions for realizing superhuman intelligence and performance, delivering sustainable competitive advantages in the age of AI.

HBR Reprint H05TFV

Originally published on HBR.org August 24, 2020

Nada R. Sanders *is a Distinguished Professor at D'Amore McKim School of Business at Northeastern University and a fellow of the Decision Sciences Institute. She is a coauthor, with John D. Wood, of* The Humachine: Humankind, Machines, and the Future of Enterprise *(Routledge, 2019).* ***John D. Wood*** *is a partner at the law firm Green Klein & Wood, where he advises commercial property owners. He is a coauthor, with Nada R. Sanders, of* The Humachine: Humankind, Machines, and the Future of Enterprise.

ORIGINALLY PUBLISHED JANUARY–FEBRUARY 2021

When Machine Learning Goes Off the Rails

A guide to managing the risks

→ by BORIS BABIC, I. GLENN COHEN, THEODOROS EVGENIOU, and SARA GERKE

WHAT HAPPENS WHEN machine learning—computer programs that absorb new information and then change how they make decisions—leads to investment losses, biased hiring or lending, or car accidents? Should businesses allow their smart products and services to autonomously evolve, or should they "lock" their algorithms and periodically update them? If firms choose to do the latter, when and how often should those updates happen? And how should companies evaluate and mitigate the risks posed by those and other choices?

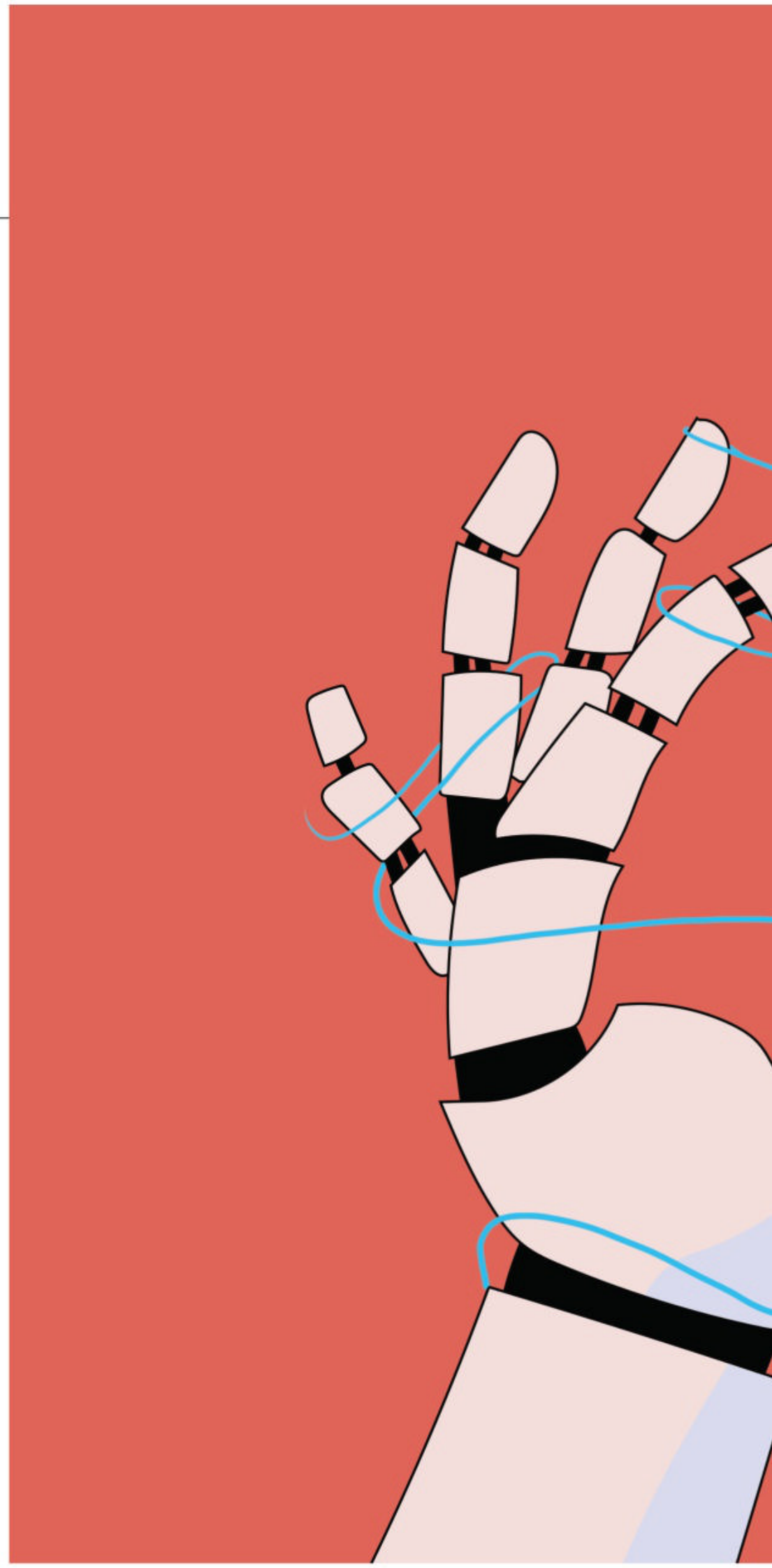
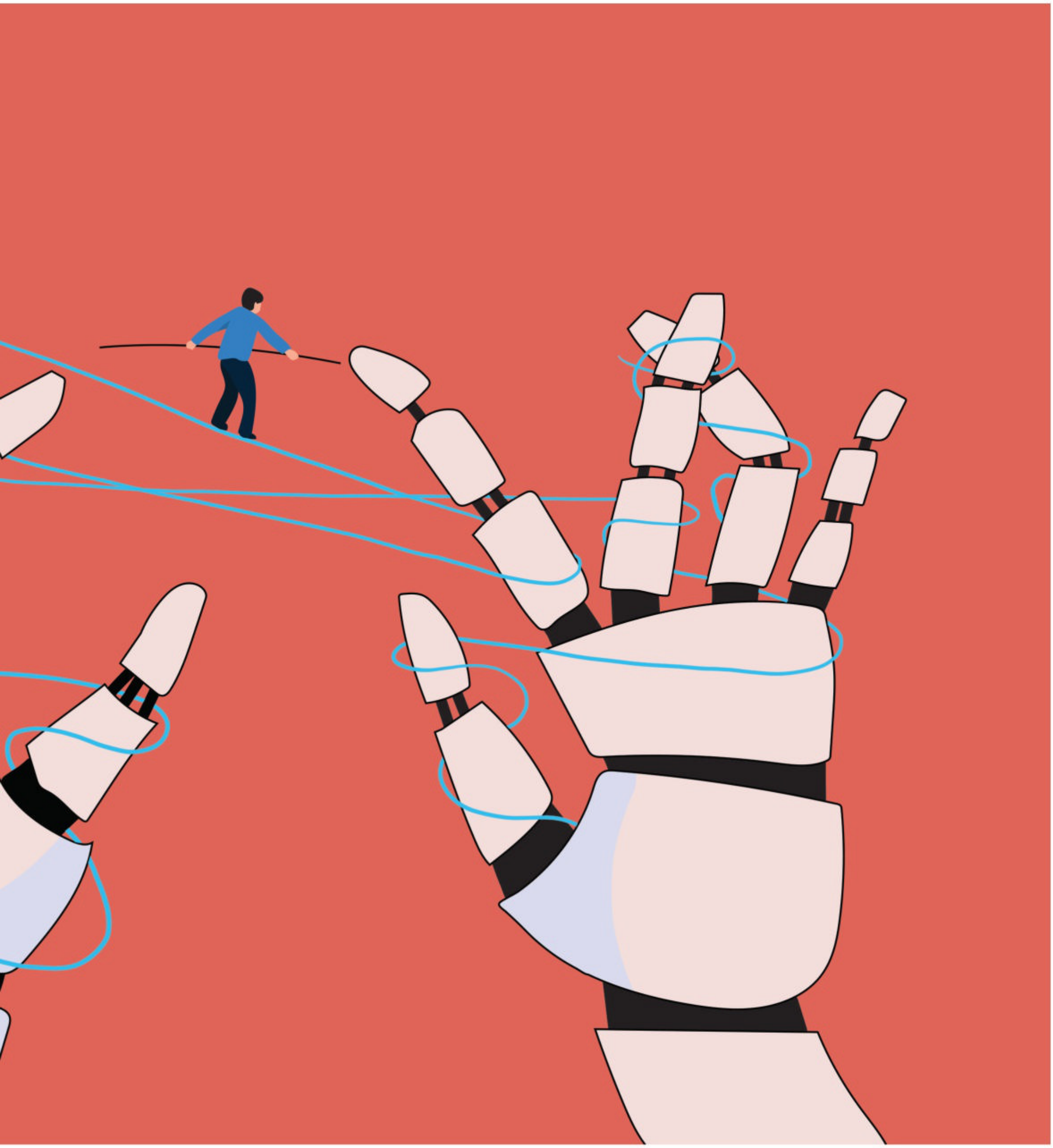


Illustration by **MICHELLE D'URBANO**



Across the business world, as machine-learning-based artificial intelligence permeates more and more offerings and processes, executives and boards must be prepared to answer such questions. In this article, which draws on our work in health care law, ethics, regulation, and machine learning, we introduce key concepts for understanding and managing the potential downside of this advanced technology.

What Makes Machine Learning Risky

The big difference between machine learning and the digital technologies that preceded it is the ability to independently make increasingly complex decisions—such as which financial products to trade, how vehicles react to obstacles, and whether a patient has a disease—and continuously adapt in response to new data. But these algorithms don't always work smoothly. They don't always make ethical or accurate choices. There are three fundamental reasons for this.

One is simply that the algorithms typically rely on the *probability* that someone will, say, default on a loan or have a disease. Because they make so many predictions, it's likely that *some* will be wrong, just because there's always a chance that they'll be off. The likelihood of errors depends on a lot of factors, including the amount and quality of the data used to train the algorithms, the specific type of machine-learning method chosen (for example, deep learning, which uses complex mathematical models, versus classification trees that rely on decision rules), and whether the system uses only *explainable algorithms* (meaning humans can describe how they arrived at their decisions), which may not allow it to maximize accuracy.

Second, the environment in which machine learning operates may itself evolve or differ from what the algorithms were developed to face. While this can happen in many ways, two of the most frequent are *concept drift* and *covariate shift*.

With the former the relationship between the inputs the system uses and its outputs isn't stable over time or may be misspecified. Consider a machine-learning algorithm for stock trading. If it has been trained using data only from a period of low market volatility and high economic growth, it may not perform well when the economy enters a recession or experiences turmoil—say, during a crisis like the Covid-19 pandemic. As the market changes, the relationship between the inputs and outputs—for example, between how leveraged a company is and its stock returns—also may change. Similar misalignment may happen with credit-scoring models at different points in the business cycle.

In medicine, an example of concept drift is when a machine-learning-based diagnostic system that uses skin images as inputs in detecting skin cancers fails to make correct diagnoses because the relationship between, say, the color of someone's skin (which may vary with race or sun exposure) and the diagnosis decision hasn't been adequately captured. Such information often is not even available in electronic health records used to train the machine-learning model.

Covariate shifts occur when the data fed into an algorithm during its use differs from the data that trained it. This can happen even if the patterns the algorithm learned are stable and there's no concept drift. For example, a medical device company may develop its machine-learning-based system using data from large urban hospitals. But once the device is out in the market, the medical data fed into the system by care providers in rural areas may not look like the development data. The urban hospitals might have a higher concentration of patients from certain sociodemographic groups who have underlying medical conditions not commonly seen in rural hospitals. Such disparities may be discovered only when the device makes more errors while out in the market than it did during testing. Given the diversity of markets and the pace at which they're changing, it's becoming increasingly challenging to foresee what will happen in the environment that systems operate in, and no amount of data can capture all the nuances that occur in the real world.

The third reason machine learning can make inaccurate decisions has to do with the complexity of the overall systems it's embedded in. Consider a device used to diagnose a disease on the basis of images that doctors input—such as IDx-DR, which identifies eye disorders like diabetic retinopathy and macular edema and was the first autonomous machine-learning-based medical device authorized for use by the U.S. Food and Drug Administration. The quality of any diagnosis depends on how clear the images provided are, the specific algorithm used by the device, the data that

How should we program an autonomous car to value the lives of three elderly people against, say, the life of one middle-aged person?

algorithm was trained with, whether the doctor inputting the images received appropriate instruction, and so on. With so many parameters, it's difficult to assess whether and why such a device may have made a mistake, let alone be certain about its behavior.

But inaccurate decisions are not the only risks with machine learning. Let's look now at two other categories: agency risk and moral risk.

Agency Risk

The imperfections of machine learning raise another important challenge: risks stemming from things that aren't under the control of a specific business or user.

Ordinarily, it's possible to draw on reliable evidence to reconstruct the circumstances that led to an accident. As a result, when one occurs, executives can at least get helpful estimates of the extent of their company's potential liability. But because machine learning is typically embedded within a complex system, it will often be unclear what led to a breakdown—which party, or "agent" (for example, the algorithm developer, the system deployer, or a partner), was responsible for an error and whether there was an issue with the algorithm, with some data fed to it by the user, or with the data used to train it, which may have come from multiple third-party vendors. Environmental change and the probabilistic nature of machine learning make it even harder to attribute responsibility to a particular agent. In fact, accidents or unlawful decisions can occur even without negligence on anyone's part—as there is simply always the possibility of an inaccurate decision.

Executives need to know when their companies are likely to face liability under current law, which may itself also evolve. Consider the medical context. Courts have historically viewed doctors as the final decision-makers and have therefore been hesitant to apply product liability to medical software makers. However, this may change as more black-box or autonomous systems make diagnoses and recommendations without the involvement of (or with much weaker involvement by) physicians in clinics. What will happen, for example, if a machine-learning system recommends a nonstandard treatment for a patient (like a much higher drug dosage than usual) and regulation evolves in such a way that the doctor would most likely be held liable for any harm only if he or she did not follow the system's recommendation? Such regulatory changes may shift liability risks from doctors to the developers of the machine-learning-enabled medical devices, the data providers involved in developing the algorithms, or the companies involved in installing and deploying the algorithms.

Moral Risk

Products and services that make decisions autonomously will also need to resolve ethical dilemmas—a requirement that raises additional risks and regulatory and product development challenges. Scholars have now begun to frame these challenges as problems of *responsible algorithm design*. They include the puzzle of how to automate moral reasoning. Should Tesla, for example, program its cars to think in

Idea in Brief

THE PROBLEM
Offerings that rely on machine learning are proliferating, raising all sorts of new risks for companies that develop and use them or supply data to train them. That's because such systems don't always make ethical or accurate choices.

THE CAUSES
First, the systems often make decisions based on probabilities. Second, their environments may evolve in an unanticipated way. Third, their complexity makes it difficult to determine whether or why they made a mistake.

THE SOLUTIONS
Executives must decide whether to let a system continuously evolve or introduce locked versions at intervals. In addition, they should test the offering appropriately before and after it is rolled out and monitor it constantly once it's on the market.

Executives need to think of machine learning as a living entity, not an inanimate technology.

utilitarian cost-benefit terms or Kantian ones, where certain values cannot be traded off regardless of benefits? Even if the answer is utilitarian, quantification is extremely difficult: How should we program a car to value the lives of three elderly people against, say, the life of one middle-aged person? How should businesses balance trade-offs among, say, privacy, fairness, accuracy, and security? Can all those kinds of risks be avoided?

Moral risks also include biases related to demographic groups. For example, facial-recognition algorithms have a difficult time identifying people of color; skin-lesion-classification systems appear to have unequal accuracy across race; recidivism-prediction instruments give Blacks and Hispanics falsely high ratings, and credit-scoring systems give them unjustly low ones. With many widespread commercial uses, machine-learning systems may be deemed unfair to a certain group on some dimensions.

The problem is compounded by the multiple and possibly mutually incompatible ways to define fairness and encode it in algorithms. A lending algorithm can be calibrated—meaning that its decisions are independent of group identity after controlling for risk level—while still disproportionately denying loans to creditworthy minorities. As a result, a company can find itself in a "damned if you do, damned if you don't" situation. If it uses algorithms to decide who receives a loan, it may have difficulty avoiding charges that it's discriminating against some groups according to one of the definitions of fairness. Different cultures may also accept different definitions and ethical trade-offs—a problem for products with global markets. A February 2020 European Commission white paper on AI points to these challenges: It calls for the development of AI with "European values," but will such AI be easily exported to regions with different values?

Finally, all these problems can also be caused by model instability. This is a situation where inputs that are close to one another lead to decisions that are far apart. Unstable algorithms are likely to treat very similar people very differently—and possibly unfairly.

All these considerations, of course, don't mean that we should avoid machine learning altogether. Instead, executives need to embrace the opportunities it creates while making sure they properly address the risks.

To Lock or Not to Lock?

If leaders decide to employ machine learning, a key next question is: Should the company allow it to continuously evolve or instead introduce only tested and locked versions at intervals? Would the latter choice mitigate the risks just described?

This problem is familiar to the medical world. The FDA has so far typically approved only "software as a medical device" (software that can perform its medical functions without hardware) whose algorithms are locked. The reasoning: The agency has not wanted to permit the use of devices whose diagnostic procedures or treatment pathways keep changing in ways it doesn't understand. But as the FDA and other regulators are now realizing, locking the algorithms may be just as risky, because it doesn't necessarily remove the following dangers:

Inaccurate decisions. Locking doesn't alter the fact that machine-learning algorithms typically base decisions on estimated probabilities. Moreover, while the input of more data usually leads to better performance, it doesn't always, and the amount of improvement can vary; improvements in unlocked algorithms may be greater or smaller for different systems and with different volumes of data. Though it's difficult to understand how the accuracy (or inaccuracy) of decisions may change when an algorithm is unlocked, it's important to try.

Environmental changes. It also matters whether and how the environment in which the system makes decisions is evolving. For example, car autopilots operate in environments that are constantly altered by the behavior of other drivers. Pricing, credit scoring, and trading systems may face a shifting market regime whenever the business cycle enters a new phase. The challenge is ensuring that the machine-learning system and the environment coevolve in a way that lets the system make appropriate decisions.

Agency risks. Locking an algorithm doesn't eliminate the complexity of the system in which it's embedded. For example, errors caused by using inferior data from third-party vendors to train the algorithm or by differences in skills across users can still occur. Liability can still be challenging to assign across data providers, algorithm developers, deployers, and users.

Moral risks. A locked system may preserve imperfections or biases un-

known to its creators. When analyzing mammograms for signs of breast cancer, a locked algorithm would be unable to learn from new subpopulations to which it is applied. Since average breast density can differ by race, this could lead to misdiagnoses if the system screens people from a demographic group that was underrepresented in the training data. Similarly, a credit-scoring algorithm trained on a socioeconomically segregated subset of the population can discriminate against certain borrowers in much the same way that the illegal practice of redlining does. We want algorithms to correct for such problems as soon as possible by updating themselves as they "observe" more data from subpopulations that may not have been well represented or even identified before. Conversely, devices whose machine-learning systems are not locked could harm one or more groups over time if they're evolving by using mostly data from a different group. What's more, identifying the point at which the device gets comparatively worse at treating one group can be hard.

A Tool Kit for Executives

So how should executives manage the existing and emerging risks of machine learning? Developing appropriate processes, increasing the savviness of management and the board, asking the right questions, and adopting the correct mental frame are important steps.

Treat machine learning as if it's human. Executives need to think of machine learning as a living entity, not an inanimate technology. Just as cognitive testing of employees won't reveal how they'll do when added to a preexisting team in a business, laboratory testing cannot predict the performance of machine-learning systems in the real world. Executives should demand a full analysis of how employees, customers, or other users will apply these systems and react to their decisions. Even when not required to do so by regulators, companies may want to subject their new machine-learning-based products to randomized controlled trials to ensure their safety, efficacy, and fairness prior to rollout. But they may also want to analyze products' decisions in the actual market, where there are various types of users, to see whether the quality of decisions differs across them. In addition, companies should compare the quality of decisions made by the algorithms with those made in the same situations *without* employing them. Before deploying products at scale, especially but not only those that haven't undergone randomized controlled trials, companies should consider testing them in limited markets to get a better idea of their accuracy and behavior when various factors are at play—for instance, when users don't have equal expertise, the data from sources varies, or the environment changes. Failures in real-world settings signal the need to improve or retire algorithms.

Think like a regulator and certify first. Businesses should develop plans for certifying machine-learning offerings before they go to market. The practices of regulators offer a good road map. In 2019, for example, the FDA published a discussion paper that proposed a new regulatory framework for modifications to machine-learning-based software as a medical device. It laid out an approach that would allow such software to continuously improve while maintaining the safety of patients, which included a complete assessment of the company—or team—developing the software to ensure it had a culture of organizational excellence and high quality that would lead it to regularly test its machine-learning devices. If companies don't adopt such certification processes, they may expose themselves to liability—for example, for performing insufficient due diligence.

Many start-ups provide services to certify that products and processes don't suffer from bias, prejudice, stereotypes, unfairness, and other pitfalls. Professional organizations, such as the Institute of Electrical and Electronics Engineers and the International Organization for Standardization, are also developing standards for such certification, while companies like Google offer AI ethics services that examine multiple dimensions, ranging from the data used to train systems, to their behavior, to their impact on well-being. Companies might need to develop similar frameworks of their own.

Monitor continuously. As machine-learning-based products and services and the environments they operate in evolve, companies may find that their technologies don't perform as initially intended. It is therefore important that they set up ways to check that these technologies behave within appropriate limits. Other sectors can serve as models. The FDA's Sentinel Initiative draws from disparate data sources, such as electronic health records, to monitor the safety of medical products and can force them to be withdrawn if they don't

pass muster. In many ways companies' monitoring programs may be similar to the preventive maintenance tools and processes currently used by manufacturing or energy companies or in cybersecurity. For example, firms might conduct so-called adversarial attacks on AI like those used to routinely test the strength of IT systems' defenses.

Ask the right questions. Executives and regulators need to delve into the following:

Accuracy and competitiveness. How much is the performance of the machine-learning-based system likely to improve with the volume of new data from its use if we don't lock the algorithm? What will such improvements mean for the business? To what extent will consumers understand the benefits and drawbacks of locked versus unlocked systems?

Biases. What data was used to train the algorithm? How representative is it of the population on which the algorithm will ultimately operate? Can we predict whether an unlocked algorithm will produce less-biased results than a locked one if we allow it to learn over time? Do the algorithm's errors affect minorities or other groups in particular? Can a continuous monitoring approach establish "guardrails" that stop the algorithm from becoming discriminatory?

The environment. How will the environment in which the offering is used change over time? Are there conditions under which machine learning should not be allowed to make decisions, and if so, what are they? How can we ensure that the offering's behavior evolves appropriately given how the environment itself is changing? When should we withdraw our offering because the gap between the environment and our offering's behavior has become too big? What are the boundaries of the environment within which our offering can adapt and operate? How robust and safe are our machine-learning systems throughout their life cycles?

Agency. On which third-party components, including data sources, does the behavior of our machine-learning algorithms depend? How much does it vary when they're used by different types of people—for example, less-skilled ones? What products or services of other organizations use our data or machine-learning algorithms, possibly exposing us to liability? Should we allow other organizations to use machine-learning algorithms that we develop?

Develop principles that address your business risks. Businesses will need to establish their own guidelines, including ethical ones, to manage these new risks—as some companies, like Google and Microsoft, have already done. Such guidelines often need to be quite specific (for example, about what definitions of fairness are adopted) to be useful and must be tailored to the risks in question. If you're using machine learning to make hiring decisions, it would be good to have a model that is simple, fair, and transparent. If you're using machine learning to forecast the prices of commodity futures contracts, you may care less about those values and more about the maximum potential financial loss allowed for any decision that machine learning makes.

Luckily, the journey to develop and implement principles doesn't need to be a lonely one. Executives have a lot to learn from the multiyear efforts of institutions such as the OECD, which developed the first intergovernmental AI principles (adopted in 2019 by many countries). The OECD principles promote innovative, trustworthy, and responsibly transparent AI that respects human rights, the rule of law, diversity, and democratic values, and that drives inclusive growth, sustainable development, and well-being. They also emphasize the robustness, safety, security, and continuous risk management of AI systems throughout their life cycles.

The OECD's recently launched AI Policy Observatory provides further useful resources, such as a comprehensive compilation of AI policies around the world.

MACHINE LEARNING HAS tremendous potential. But as this technology, along with other forms of AI, is woven into our economic and social fabric, the risks it poses will increase. For businesses, mitigating them may prove as important as—and possibly more critical than—managing the adoption of machine learning itself. If companies don't establish appropriate practices to address these new risks, they're likely to have trouble gaining traction in the marketplace.

HBR Reprint R2101F

Boris Babic *is an assistant professor of philosophy and of statistics at The University of Toronto.* ***I. Glenn Cohen*** *is a deputy dean, professor of law, and faculty director of the Petrie-Flom Center for Health Law Policy, Biotechnology, and Bioethics at Harvard Law School.* ***Theodoros Evgeniou*** *is a professor of decision sciences and technology management at INSEAD.* ***Sara Gerke*** *is an assistant professor of law at Penn State Dickinson Law.*

Gongos, Inc.
Congratulates
Vex Award Winners

Do you know what these brands have in common?

Peloton
John Deere
Warby Parker
Louis Vuitton
Cricket Wireless

Based on Gongos' patent-pending Value Exchange (Vex) Model, these brands are delivering on outcomes that matter most to customers, and as a result, stand to increase their Customer Lifetime Value.

Discover your Vex Curve™—the first-ever opportunity to see the proven link between your business outcomes and customer goals.

Visit vex.gongos.com to explore your brand's growth potential among the best in class.

gongos | WE MAKE CUSTOMER CENTRICITY HAPPEN

Quick Takes

MIRAGEC/GETTY IMAGES

1. What Will Happen When Your Company's Algorithms Go Wrong?

→ by ROMAN V. YAMPOLSKIY

WHEN YOU'RE READY to incorporate AI technologies in your business, the analysis you should perform is this: What can possibly go wrong? What is our product or service expected to do? What happens if it fails to do so? Do we have a damage mitigation plan? Consider the embarrassing situation that Microsoft found itself in with its Tay chatbot fiasco, where internet trolls exploited vulnerabilities in the bot's code, feeding it racist, homophobic, and sexist content that millions read on social media.

Accidents, including deadly ones, caused by software or industrial robots can be traced to the early days of such technology, but they are not necessarily caused by the systems themselves. AI failures, on the other hand, are directly related to the mistakes produced by the intelligence such systems are designed to exhibit. We can broadly classify such failures into "mistakes made during the learning phase" and "mistakes made during the performance phase." A system can fail to learn what its designers want it to learn and might instead learn a different but correlated function.

A frequently cited example is a computer vision system that the U.S. Army had hoped to use to automatically detect camouflaged enemy tanks. The system was supposed to classify pictures of tanks but instead learned to distinguish the backgrounds of such images. Other examples include problems caused by poorly

The frequency and seriousness of AI failures will steadily increase as AIs become more capable.

designed functions that would reward AIs for only partially desirable behaviors, such as pausing a game to avoid losing or repeatedly touching a soccer ball to get credit for possession.

It can help to look at some recent examples of AI failure to better understand what problems are most likely to arise and what you can do to prevent them—or at least to clean up quickly after a failure. Consider these examples of AI failures from the past few years:

2015: An automated email-reply generator created inappropriate responses, such as writing "I love you" to a business colleague.

2015: A robot for grabbing auto parts grabbed and killed a man.

2015: Image-tagging software classified Black people as gorillas.

2015: Medical AI classified patients with asthma as having a lower risk of dying of pneumonia.

2015: Adult-content filtering software failed to remove inappropriate content, exposing children to violent and sexual content.

2016: AI designed to predict recidivism acted racist.

2016: An AI agent exploited a reward signal to win a game without actually completing the game.

2016: Video game NPCs (nonplayer characters, or any character that is not controlled by a human player) designed unauthorized superweapons.

2016: AI judged a beauty contest and rated dark-skinned contestants lower.

2016: A mall security robot collided with and injured a child.

2016: The AI AlphaGo lost to a human in a world-championship-level game of Go.

2016: A self-driving car had a deadly accident.

And every day, consumers experience more-common shortcomings of AI: Spam filters block important emails, GPS provides faulty directions, machine translations corrupt the meaning of phrases, autocorrect replaces a desired word with a wrong one, biometric systems misrecognize people, transcription software fails to capture what is being said. Overall, it is harder to find examples of AIs that *don't* fail.

Analyzing the list of AI failures above, we can arrive at a simple generalization: An AI designed to do X will eventually fail to do X. While it may seem trivial, it is a powerful generalization tool that can be used to predict future failures of AIs. For example, looking at cutting-edge current and future AIs, we can predict the following:

- AI doctors will misdiagnose some patients in a way a real doctor would not.
- Video-description software will misunderstand movie plots.
- Software for generating jokes will occasionally fail to make them funny.
- Sarcasm-detection software will confuse sarcastic and sincere statements.
- Employee-screening software will be systematically biased and thus hire low performers.
- The Mars robot-explorer will misjudge its environment and fall into a crater.
- Tax-preparation software will miss important deductions or make inappropriate ones.

What should you learn from the above examples and analysis? Failures will happen! It's inevitable. But, you can still put best practices in place, such as:

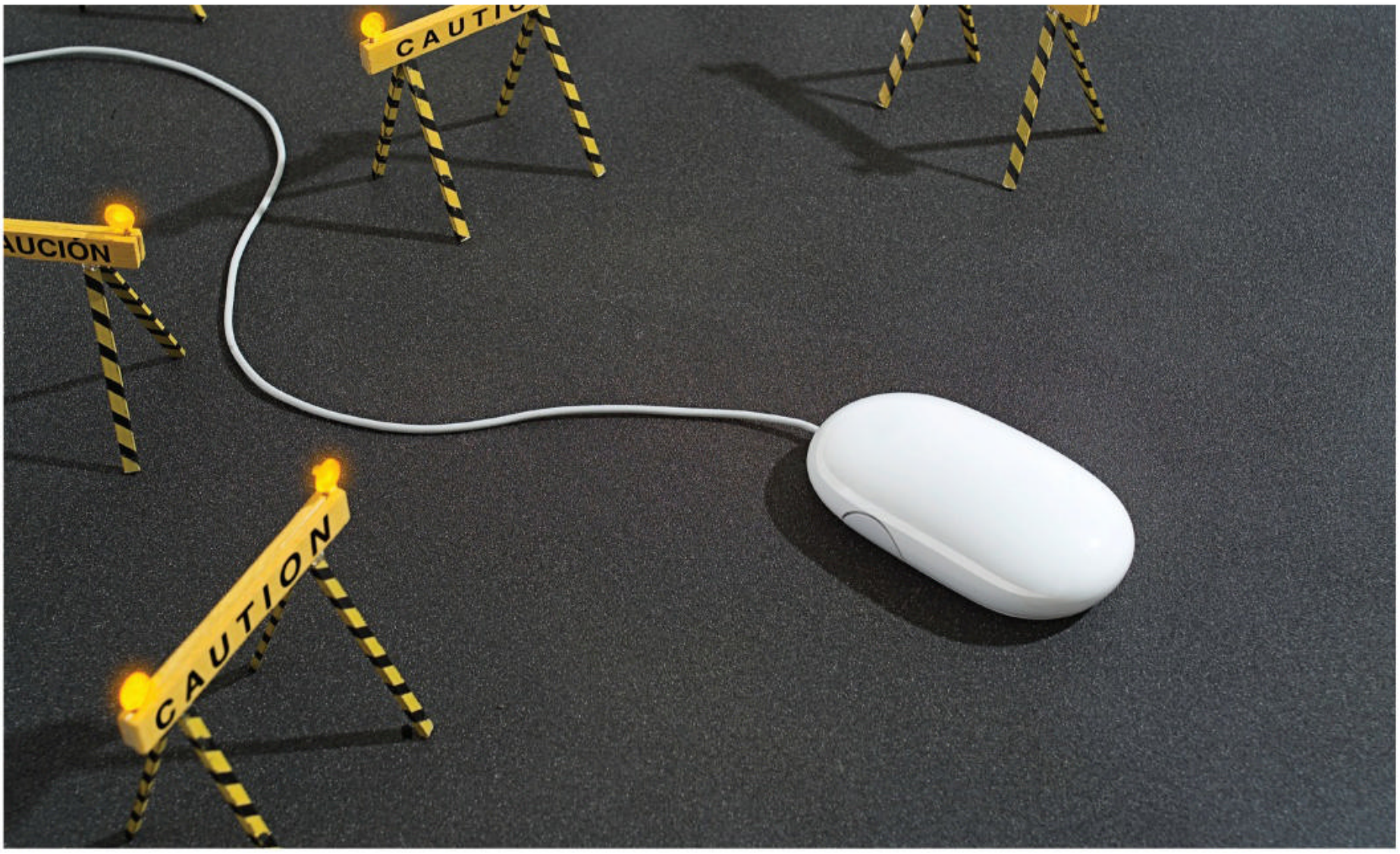
- Controlling user input to the system and limiting learning to verified data inputs
- Checking for racial, gender, age, and other common biases in your algorithms
- Explicitly analyzing how your software can fail, and then providing a safety mechanism for each possible failure
- Having a less "smart" backup product or service available
- Having a communications plan in place to address the media in case of an embarrassing failure (Hint: Start with an apology.)

I predict that both the frequency and seriousness of AI failures will steadily increase as AIs become more capable. The failures of today's narrow-domain AIs are just the tip of the iceberg. Once we develop general AI capable of cross-domain performance, embarrassment will be the least of our concerns.

HBR Reprint H03L3S

Originally published on HBR.org April 10, 2017

Roman V. Yampolskiy *is a tenured associate professor in the Department of Computer Engineering and Computer Science at the Speed School of Engineering, University of Louisville. He is the founding and current director of the university's cybersecurity lab and an author of many books, including* Artificial Superintelligence: A Futuristic Approach *(Chapman & Hall, 2017). Follow him on Twitter @romanyam.*



C.J. BURTON/GETTY IMAGES

2. What People Hate About Being Managed by Algorithms, According to a Study of Uber Drivers

→ by MAREIKE MÖHLMANN and OLA HENFRIDSSON

COMPANIES ARE increasingly using algorithms to manage their remote workforces. Called *algorithmic management,* this approach has been most widely adopted in gig economy companies. For example, the ride-hailing company Uber substantially increases its efficiency by managing some 3 million workers with an app that instructs drivers which passengers to pick up and which routes to take.

Algorithmic management is frustrating to workers, and their resentment can lead them to behave subversively.

Being managed in this way offers some benefit to self-employed workers as well: For example, Uber drivers are free to decide when and for how long they would like to work and which area they would like to serve. However, our research reveals that algorithmic management is also frustrating to workers, and their resentment can lead them to behave subversively, which could cause real harm to their companies. Our research also suggests some ways that companies can mitigate these concerns while still taking advantage of the benefits of management by algorithm.

Together with Lior Zalmanson (Tel Aviv University) and Robert W. Gregory (University of Virginia), we conducted a multimethod study of Uber drivers in New York and London. We collected data by informally and formally interviewing 34 drivers, observing drivers in action, analyzing more than 1,000 online forum posts, and reviewing media coverage of Uber in several waves between December 2015 and September 2018.

We found that Uber drivers have three areas of consistent complaints about working "for" algorithms, concerns we've also seen in other companies using algorithmic management:

Constant surveillance. As soon as they log on to the Uber app, drivers are watched and scrutinized by the platform's algorithms. The app tracks their GPS location, speed, and acceptance rate of customer requests. It instructs them about which riders to pick up where and how to get to the riders' destinations. If the drivers diverge from the app's instructions, they can be penalized or even banned from the platform. Regardless of whether the attention comes through an app or in person, we know that scrutiny of work can reduce productivity.

From our study we found that drivers found performance evaluations in the form of customer ratings particularly frustrating. This may be because they amplify Uber drivers' negative feelings about constant surveillance with an additional layer of technology-mediated attention.

Little transparency. Although the app is learning a lot about them, Uber drivers find it frustrating how little they know about the app. They find the lack of transparency of the underlying logic of the complex algorithms frustrating, believing it to be an unfair system that manipulates them subtly without their knowledge or consent. (Indeed, Uber has previously admitted to drawing on insights from behavioral science to nudge drivers to work longer hours.)

Uber drivers, as well as other gig economy workers such as courier and delivery workers at Postmates and Deliveroo, are demanding more transparency about the allocation of jobs, the compilation of their ratings, and their payment structure. However, companies such as Uber argue they can't reveal the secret recipe of their algorithms to competitors. Furthermore, recent advances in AI and machine learning mean that algorithms can now learn and dynamically adjust to any given environment, allowing for the automation of more-sophisticated tasks (such as managing the workforce). But the more sophisticated these algorithms get, the more opaque they are, even to their creators.

Dehumanization. Drivers at Uber report feeling equally lonely, isolated, and dehumanized. They don't have colleagues to socialize with or a team or community to be part of. They lack the opportunity to build a personal relationship with a supervisor. Those on crowdworking platforms like Amazon Mechanical Turk have raised similar complaints as they conduct "microtasks" such as classifying content or participating in surveys.

Drivers have responded to their various frustrations with these algorithms by identifying clever ways to work around them. For instance, one driver in the Uberpeople.net forum wrote: "Play the system, don't let it play you. We all know that these companies like to offer better incentives to drivers that miss some time. So, drive Uber for one week, Juno next, Lyft third and etc. I switch between Uber/Juno weekly."

They are also angry enough—and feel disempowered enough—that they are finding creative ways to make their displeasure known. For instance, drivers are gaming the system by artificially causing surge pricing. They are also getting political. Especially in the gig economy, drivers of ride-hailing services and couriers seek to compensate for the social isolation they experience in their everyday routine by actively

Getting workers actively participating in discussions about the design of algorithm-driven systems would create more engaged and supportive workforces.

engaging in online communities, but companies themselves aren't involved in those platforms. Instead, more-adversarial union-type organizations have sprung up as drivers or couriers look to support one another, such as the Seattle-based workers' rights organization Working Washington, which drew together couriers delivering for Postmates, DoorDash, and other on-demand services.

To address these challenges and mitigate their risks to the business—and, not least, to implement ethical practices—we suggest that companies who manage all or part of their workforce through algorithms:

1. Share information. In theory, algorithmic management can increase transparency, because even learning algorithms that are used to manage workers reflect a set of rules and procedures that comply with the strategic goals of upper management. It may not be possible to share the algorithm itself with workers, but company leadership can and should share with them the data and goals that informed it.

2. Invite feedback. To counterbalance the unidirectional commands that the algorithm hands down to drivers, companies should find ways to democratically include them in decision making, for example, by involving them in committees or councils that discuss and negotiate work related to internal regulations. Getting workers actively participating in discussions about the design of algorithm-driven systems would create more engaged and supportive workforces.

3. Build in human contact. People need people. Organizations should develop formal, supportive communities where workers feel like members and can make social connections. Adding a human element to the way people are managed will help workers feel less like they are being treated as machines. For instance, some of the drivers in our study spoke fondly of New York ride-hailing firm Juno (acquired by Gett in 2017), which, early in its existence, employed an extensive human customer support system that eagerly helped drivers with questions or problems.

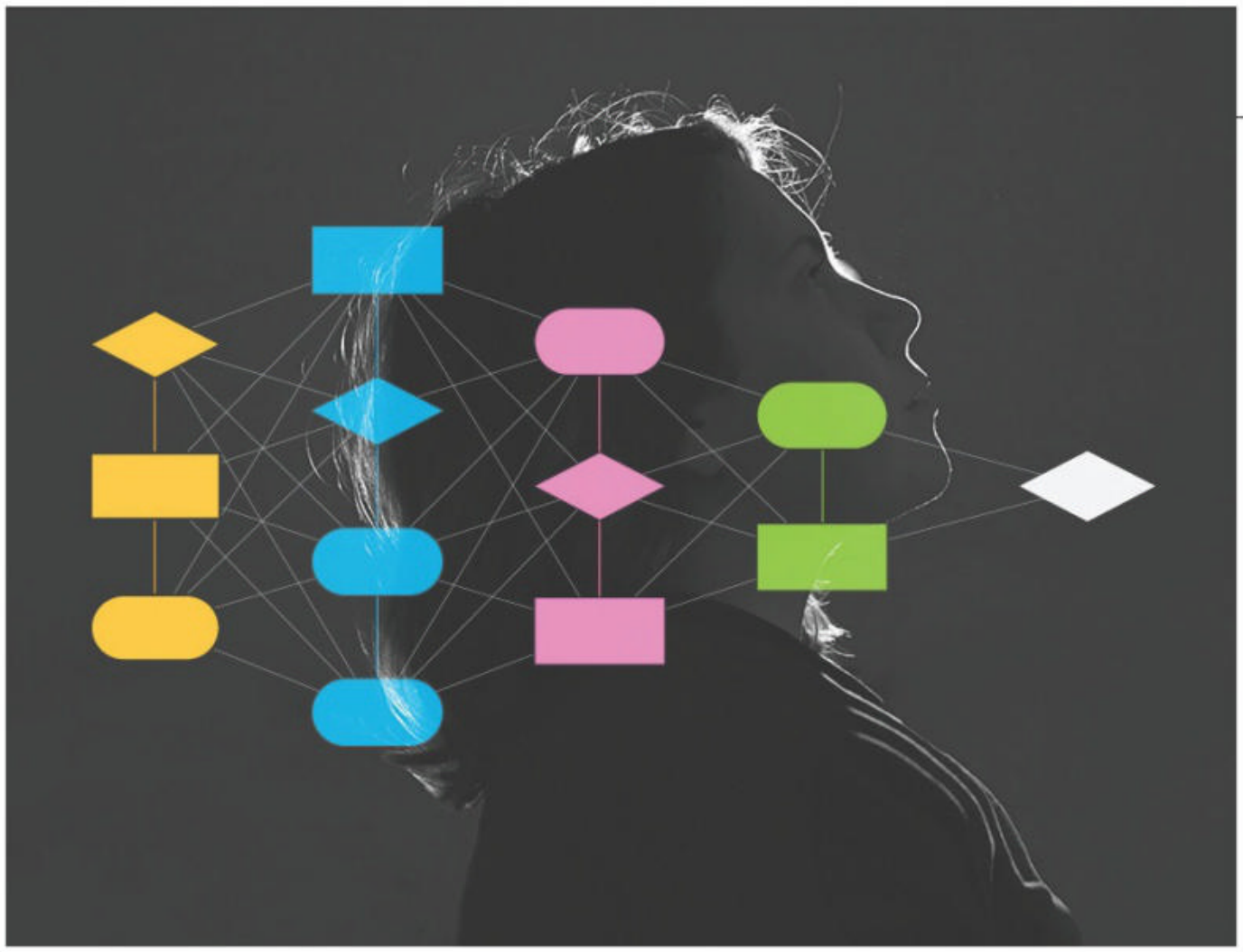
4. Build trust. Implementing benefits that improve workers' welfare, such as providing financial support in case of illness, or better sick pay or maternity leave, may be a first step to humanizing the company and mitigating the anger of employees who are managed by faceless algorithms.

Regulators across the globe are already seeking to implement regulations around some of these recommendations to benefit an algorithmically managed workforce. For instance, in 2017 it was ruled that Uber will need to pay UK-based drivers a minimum wage and provide sick and holiday pay. Given the rapid pace of technological progress, and the tempting economic benefits for companies, we believe that algorithmic management will become more common in the coming years. As more companies manage their labor force in this way—and as they incur the anger of the workforce that makes their core offerings possible—it becomes that much more incumbent upon them to take some of these steps on their own.

HBR Reprint H054VX

Originally published on HBR.org August 30, 2019

***Mareike Möhlmann** is an assistant professor at Bentley University. Previously she worked as an assistant professor at Warwick Business School, a teaching fellow/visiting assistant professor at the London School of Economics, and a postdoctoral researcher at the New York University Stern School of Business. **Ola Henfridsson** is a professor of business technology at the University of Miami Business School. He previously served as professor and head of the Information Systems group at Warwick Business School in the UK.*



HBR STAFF/KLARA KULOVEC/EYEEM/GETTY IMAGES

3. Why Business Leaders Need to Understand Their Algorithms

→ by MIKE WALSH

ONE OF THE biggest sources of anxiety about AI is not that it will turn against us but that we simply cannot understand how it works. The solution to rogue systems that discriminate against women in credit applications, or make racist recommendations in criminal sentencing, or reduce the number of Black patients identified as needing extra medical care, might seem to be "explainable AI." But sometimes, what's just as important as knowing "why" an algorithm made a decision is being able to ask "what" it was being optimized for in the first place?

Machine-learning algorithms are often called a black box because they resemble a closed system that takes an input and produces an output, without any explanation as to why. Knowing why is important for many industries, particularly those with fiduciary obligations like consumer finance, or in health care and education, where vulnerable lives are involved, or in military or government applications, where you need to be able to justify your decisions to the electorate.

Unfortunately, when it comes to deep-learning platforms, explainability is problematic. In many cases, the appeal of machine learning lies in its ability to find patterns that defy logic or intuition. If you could map a relationship simply enough between inputs and outputs to explain it, you probably wouldn't need machine learning in that context at all. Unlike a hand-coded system, you can't just look inside a neural network and see how it works. A neural network is composed of thousands of simulated neurons arranged in interconnected layers that each receive input and output signals that are then fed into the next layer, and so on, until a final output is reached. Even if you can interpret how a model is technically working in terms that an AI scientist could comprehend, explaining that to a "civilian decision-maker" is another problem altogether.

Deep Patient, for example, is a deep-learning platform at Mount Sinai Hospital in New York. It was trained using electronic health records from 700,000 individuals and became adept at predicting disease by discovering patterns hidden in the hospital data that provided early warnings for patients at risk of developing a wide variety of ailments, including liver cancer, without human guidance.

In the near future, you won't be able to design any kind of AI without both a team of top scientists and an equally capable team of privacy engineers and lawyers.

Then, much to everyone's surprise, Deep Patient also demonstrated an ability to predict the onset of certain psychiatric disorders like schizophrenia, which are notoriously difficult even for doctors to predict. The challenge for medical professionals in such a scenario is to balance acknowledging the efficacy and value of the system with knowing how much to trust it, given that they don't fully understand it or how it works.

Some organizations and industries are investing in the capability to audit and explain machine-learning systems. The Defense Advanced Research Projects Agency (DARPA) is currently funding a program called Explainable AI, whose goal is to interpret the deep learning that powers drones and intelligence-mining operations. Capital One, which has had its own serious issues with data breaches, created a research team dedicated to finding ways to make deep learning more explainable, as U.S. regulations require this type of company to explain decisions such as why they denied a credit card to a prospective customer.

Algorithmic regulation is likely to become more sophisticated over the next few years as the public starts to become more openly concerned about the impact of AI on their lives. For example, under the General Data Protection Regulation (GDPR), which came into effect in 2018, the European Union requires companies to be able to explain a decision made by one of its algorithms. Arguably in the near future, you won't be able to design any kind of AI without both a team of top scientists and an equally capable team of privacy engineers and lawyers.

The rationale behind algorithmic regulation is accountability. Making AI more explainable is not just about reassuring leaders that they can trust algorithmic decisions; it is also about providing recourse for people to challenge AI-based decisions. In fact, the issue of algorithmic transparency applies not only to machine learning but also to any algorithm whose inner workings are kept hidden.

Algorithms that either appear to be biased or are obscure in the way they work have already been challenged in the courts. For example, in 2014 the Houston Federation of Teachers brought a lawsuit against the Houston School District, arguing that the district's use of a secret algorithm to determine how teachers were evaluated, fired, and given bonuses was unfair. The system was developed by a private company, which classified its algorithm as a trade secret and refused to share it with teachers. Without knowing how they were being scored, teachers said, they were denied the right to challenge their terminations or evaluations. A circuit court found that the unexplainable software violated the teachers' 14th Amendment right to due process, and the case was ultimately settled in 2016, with use of the algorithm being discontinued. In the next few years, the number of such challenges is likely to rise.

However, for leaders, the most important question to ask the teams designing and building automated solutions may be not *why* they reached a particular decision but rather *what* they are being optimized for. Optimums are important. There's a classic thought experiment proposed by Swedish philosopher Nick Bostrom called the Paperclip Maximizer. It describes how an AI could end up destroying the world after being given the goal to manufacture paper clips as efficiently as possible, "with the consequence that it starts transforming first all of earth and then increasing portions of space into paper clip manufacturing facilities."

The AI in Bostrom's paper is not intrinsically evil. It was simply, in his view, given the wrong goal and no constraints. Wrong goals or optimums can cause a lot of unintended harm. For example, an AI program that set school and bus schedules in Boston was scrapped after an outcry from working parents and others who objected that it did not take into account their schedules and seemed to be focused on efficiency at the expense of education. But was it the program's fault? It was, after all, coded to look for ways to save money. However, unlike the complexities of building and interpreting an AI model, debating and deciding on what a system is optimized for is absolutely within the capability set of business leaders and boards, and so it should be.

AI is a tool that reflects our priorities as organizations and governments. It might seem cold to discuss human fatalities in automotive or workplace accidents in terms of statistics, but if we decide that an algorithmic system should be designed to minimize accidents as a whole, we have to also judge any resulting harm in the context of the system it replaces. But

in doing so, we will also have to be ready to be judged ourselves, and to justify our decisions and design principles.

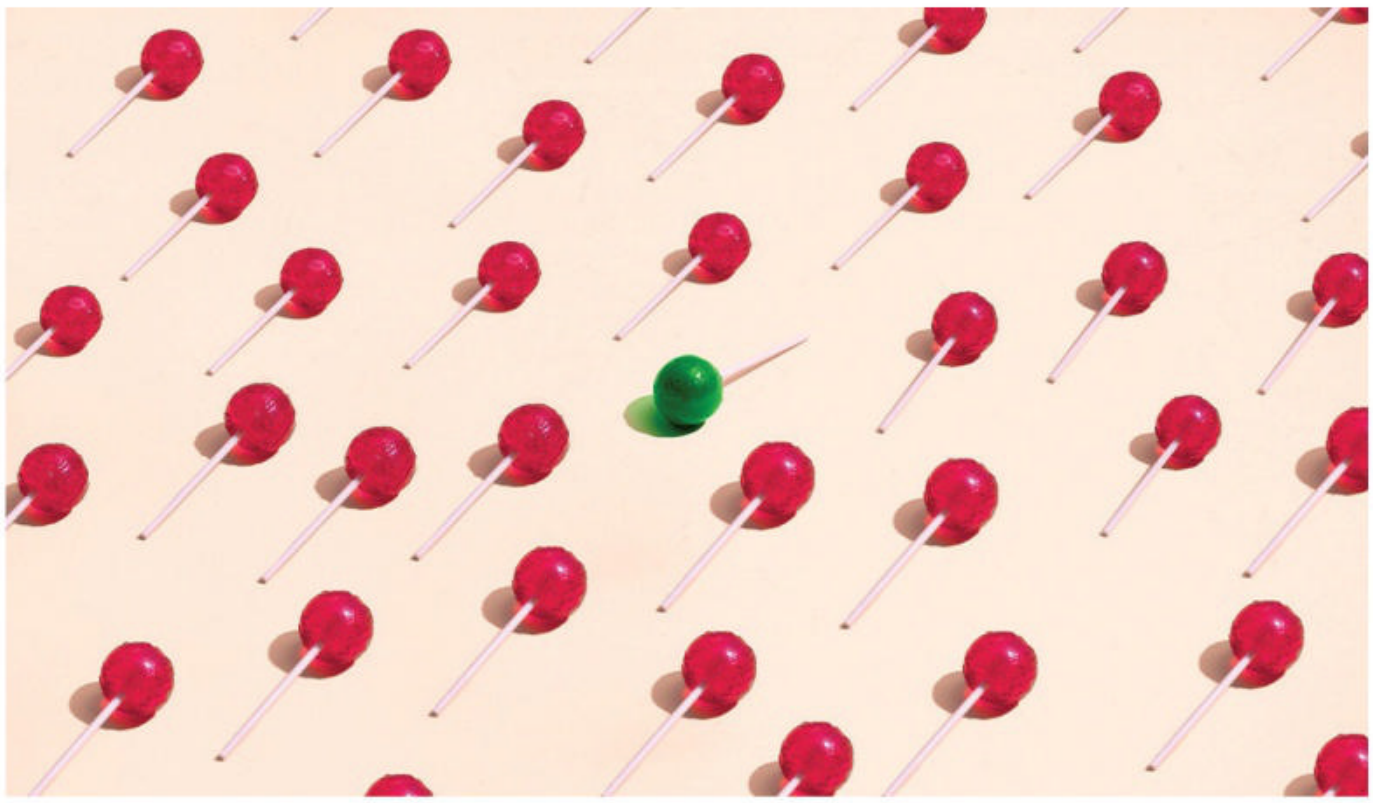
Leaders will be challenged by shareholders, customers, and regulators on what they optimize for. There will be lawsuits that require you to reveal the human decisions behind the design of your AI systems, what ethical and social concerns you took into account, the origins and methods by which you procured your training data, and how well you monitored the results of those systems for traces of bias or discrimination. Document your decisions carefully and make sure you understand, or at the very least trust, the algorithmic processes at the heart of your business.

Simply arguing that your AI platform was a black box that no one understood is unlikely to be a successful legal defense in the 21st century. It will be about as convincing as "the algorithm made me do it."

HBR Reprint H05ABP

Originally published on HBR.org November 19, 2019

Mike Walsh *is the author of* The Algorithmic Leader: How to Be Smart When Machines Are Smarter Than You *(Page Two, 2019) and CEO of Tomorrow, a global consultancy on designing companies for the 21st century.*



DANIEL GRIZELJ/GETTY IMAGES

4. If Your Company Uses AI, It Needs an Institutional Review Board

→ by REID BLACKMAN

CONVERSATIONS AROUND AI and ethics may have started as a preoccupation of activists and academics, but now—prompted by the increasing frequency of headlines of biased algorithms, black box models, and privacy violations—boards, C-suites, and data and AI leaders have realized it's an issue for which they need a strategic approach.

A solution is hiding in plain sight. Other industries have already found ways to deal with complex ethical quandaries quickly, effectively, and in a way that can be easily replicated. Instead of trying to reinvent this process, companies need to adopt and customize one of health care's greatest inventions: the institutional review board, or IRB.

Three Moves in AI Ethics Discussions

Most discussions of AI ethics follow the same flawed formula consisting of three moves, each of which is problematic from the perspective of an organization that wants to mitigate the ethical risks associated with AI.

Here's how these conversations tend to go.

First, companies move to identify AI ethics with "fairness" in AI, or sometimes more generally, "fairness, equity, and inclusion." This certainly resonates with the zeitgeist—the rise of Black Lives Matter, the anti-racist movement, and corporate

The list of ways a company can meet with ethical disaster is endless, so reducing AI ethics to issues involving fairness is a recipe for disaster.

support for diversity and inclusion measures.

Second, they move from the language of fairness to the language of "bias": biased "algorithms," as popular media puts it, or biased "models" as engineers (more accurately) call it. The examples of (allegedly) biased models are well-known, including those from Amazon, Optum Health, and Goldman Sachs.

Finally, they look for ways to address the problem as they've defined it. They discuss technical tools (whether open source, or sold by Big Tech or a start-up) for bias identification, which standardly compare a model's outputs with dozens of quantitative metrics or "definitions" of fairness found in the burgeoning academic research area of machine-learning ethics. They may also consider engaging stakeholders, especially those that comprise historically marginalized populations.

While some recent AI ethics discussions go beyond this, many of the most prominent ones don't. And the most common set of actions that practitioners actually undertake flows from these three moves. Most companies adopt a risk mitigation strategy that uses one of the aforementioned technical tools, if they're doing anything at all.

All of this should keep the stewards of brand reputation up at night, because this process has barely scratched the surface of the ethical risks that AI introduces. To understand why this is, let's take each of these moves in turn.

The first move sends you off in the wrong direction because it immediately narrows the scope. Defining "AI ethics" as "fairness in AI" is problematic because fairness issues are just a subset of ethical issues—you've just decided to ignore giant swaths of ethical risk. Most obviously there are issues relating to privacy violations (given that most of current AI is machine learning, which is often powered by people's data) and unexplainable outputs/black-box algorithms. But there are more. For example, the primary ethical risk related to AI-powered self-driving cars isn't bias or privacy, but killing and maiming. The ethical risk of facial recognition technology doesn't end when bias has been weeded out from the model (of which there are several examples); a nonbiased facial recognition software still enables surveillance by corporations and (fascist) governments. AI also requires vast amounts of energy to power the computers that are training the algorithms, which entails a shocking degree of damage to the environment. The list of ways a company can meet with ethical disaster is endless, so reducing AI ethics to issues involving fairness is a recipe for disaster.

The second move further reduces your remit: Issues of bias are a subset of issues of fairness. More specifically, issues of bias in the context of AI are issues of how different subpopulations are treated relative to others—whether goods, services, and opportunities are distributed fairly and justly. Are the job ads placed in such a way that they are just as likely to be seen by the African American population as by the white population? Are women applying for a job just as likely to have their résumés lead to an interview as a man's résumé? The problem with this approach is that issues of fairness extend beyond issues of fair distributions of goods to various subpopulations.

Most obviously, there are issues about what any individual deserves independently of how others are treated. If I'm torturing you, and you protest, it would hardly justify my actions to say, "Don't worry. I'm torturing other subpopulations at equal rates to the population of which you're a member." The entire category of human rights is about what every person deserves, independently of how others are being treated. Fairness crucially involves issues of individual desert, and a discussion (let alone a risk mitigation strategy) that leaves this out is the more perilous for it.

The final trouble for organizations arrives in the third move—identifying and adopting bias mitigation strategies and technical tools. Organizations often lean on technical tools, in particular, as their go-to (or only) meaningful instrument to ferret out bias, as measured by quantitative "definitions" of fairness found in recent computer science literature. Here, we run into a raft of failures at ethical risk mitigation:

1. Those two-dozen-plus quantitative metrics for fairness are not compatible with one another. You simply cannot be fair according to all of them at the same time. That means that an ethical judgment needs to be made: Which, if any, of these quantitative metrics of fairness

are the ethical/appropriate ones to use? Instead of bringing in lawyers, political theorists, or ethicists—all of whom have training in these kinds of complex ethical questions—these decisions are left to data scientists and engineers. But if the experts aren't in the room, you cannot expect your due diligence to have been responsibly discharged.

2. These tools generally kick in well into the development life cycle. Because they measure the output of AI models, they're used after data sets have been chosen and models have been trained—and a good deal of resources have been devoted to the product. It is then inefficient, not to mention unpopular, to go back to the drawing board if a bias problem is detected that cannot be solved in a fairly straightforward way.

3. While the search for a technical, quantitative solution to AI ethics is understandable, the truth is that many ethical issues are not reducible to quantitative metrics or KPIs. Surveillance is a problem because it destroys trust, causes anxiety, alters people's behavior, and ultimately erodes autonomy. Questions about whether people are being treated respectfully, whether a product design is manipulative or merely giving reasonable incentives, whether a decision places a burden on people that is too great to reasonably expect of them—these all require qualitative assessments.

4. These technical tools do not cover all types of bias. They do not, for instance, ferret out whether your search engine has labeled Black people "gorillas." These are cases of bias for which no technical tool exists.

5. The way these tools measure for bias is often not compatible with existing anti-discrimination law. For example, anti-discrimination law forbids companies from using variables like race and gender in their decision-making process. But what if that needs to be done to test their models for bias and thus influences the changes they make to the model in an effort to mitigate bias? That looks to be not only ethically permissible but plausibly ethically required as well.

6. As for engaging stakeholders, that is generally a good thing to do. However, aside from the logistical issues that arise, it does not by itself mitigate any ethical risks; it leaves them right in place, unless one knows how to think through stakeholder feedback. For instance, suppose your stakeholders are racist. Suppose the local norms for where you will deploy your AI encourage gender discrimination. Suppose your stakeholders disagree with one another because, in part, they have conflicting interests; stakeholders are not a monolithic group with a single perspective, after all. Stakeholder input is valuable, but you cannot programmatically derive an ethical decision from stakeholder input.

A Better Way: The AI IRB

My point here isn't that we should avoid technical tools and stakeholder outreach; they are indeed quite useful. But we need more-comprehensive ways to deal with ethical risk. Ideally, this will involve building a comprehensive AI ethical risk mitigation program that is implemented throughout an organization—admittedly a heavy lift. If a company is looking for something to do in relatively short order that can have a big impact (and will later dovetail well with the bigger risk-mitigation program), they should take their cues on ethical risk mitigation from health care and create an institutional review board (IRB).

In the United States, IRBs in medicine were introduced to mitigate the ethical risks that arose and were commonly realized in research on human subjects. Some of that unethical conduct was particularly horrific, including the Tuskegee experiments, in which doctors refrained from treating Black men with syphilis, despite penicillin being available, so that they could study the disease's unmitigated progression. More generally, the goals of an IRB include upholding the core ethical principles of respect for persons, beneficence, and justice. IRBs carry out their function by approving, denying, and suggesting changes to proposed research projects.

Comparing the kinds of ethical risks present in medicine with the kinds present in AI is useful for several reasons. First, in both instances there is the potential for harming individuals and groups of people (for example, members of a particular race or gender). Second, there exists a vast array of ethical risks that can be realized in both fields, ranging from causing physical harm and mental distress to discriminating against

Companies should take their cues on ethical risk mitigation from health care and create an institutional review board.

protected classes, invading people's privacy, and undermining people's autonomy. Third, many of the ethical risks in both instances arise from the particular applications of the technology at hand.

Applied to AI, the IRB can systematically and exhaustively identify ethical risks across the board. Just as in medical research, an AI IRB can not only approve and reject various proposals but should also make ethical risk mitigation recommendations to researchers and product developers. Moreover, a well-constituted IRB can perform the functions that the current approach cannot.

When it comes to building and maintaining an IRB, three issues loom large: membership of the board, jurisdiction, and articulating the values it will strive to achieve (or at least the nightmares it will strive to avoid).

Membership

To systematically and exhaustively identify and mitigate AI ethical risks, an AI IRB requires a diverse team of experts. You will want to place an engineer who knows the technical underpinnings of the research and/or product so that the committee can understand what is being done and what can be done from a technical perspective. Similarly, someone deeply familiar with product design is important. They speak the language of the product developers, understand "customer journeys," and can help shape ethical risk mitigation strategies that don't undermine the essential functions of the products under consideration.

You'll also want to include ethics-adjacent members, such as attorneys and privacy officers. Their knowledge of current and potential regulations, anti-discrimination law, and privacy practices are important places to look when vetting for ethical risks.

Insofar as the AI IRB has as its function the identification and mitigation of ethical risks, it would be wise to include an ethicist, for example, someone with a Ph.D. in philosophy who specializes in ethics or, say, someone with a master's degree in medical ethics. The ethicist isn't there to act as a kind of priest with superior ethical views. They're there because they have the training, knowledge, and experience related to understanding and spotting a vast array of ethical risks; the familiarity with important concepts and distinctions that aid in clear-eyed ethical deliberation; and the skill of helping groups of people objectively assess ethical issues. Importantly, this kind of risk assessment is distinct from the risks one finds in the model risk assessments created by data scientists and engineers, which tend to focus on issues relating to accuracy and data quality.

You may also find it useful to include various subject-matter experts depending on the research or product at hand. If the product is to be deployed in universities, someone deeply familiar with their operations, goals, and constituencies should be included. If the product is to be deployed in Japan, including an expert in Japanese culture may be important.

Lastly, to maintain independence and the absence of conflict of interests (for example, members looking for approval from their bosses), having at least one member unaffiliated with your organization is important (and is, incidentally, required for medical IRBs). At the same time, all members should have a sense of the business goals and necessities.

Jurisdiction

When should an AI IRB be consulted, how much power should it have, and where should it be situated in product development? In the medical community, IRBs are consulted before research begins. The reason is obvious: The IRB is consulted when one needs approval before testing on human subjects is performed. Medical IRBs are the ultimate authority. They can approve and reject proposals, as well as suggest changes to the proposal, and their decisions are final. Once an IRB has denied a proposal, another IRB cannot approve it, and the decision cannot be appealed.

The same rule should apply for an AI IRB.

Even though the harm typically occurs during deployment of the AI, not during research and product development, there's a strong case for having an AI IRB before research and/or product development begins because it's much easier—and therefore less expensive and more efficient—to change projects and products that do not yet exist. If, for instance, you only realize a significant ethical risk from a potential or probable unintended consequence of how the product was designed, you will have to either go to market with a product you know to be ethically risky or endure the costly process of reengineering the product.

While medical IRBs are granted their authority by the law, you should consider voluntarily granting that degree of power to an AI IRB: It is a tool by which great trust can be built with employees, clients, and consumers. That is particularly true if your organization is transparent about the operations—even if not the exact decisions—of the IRB. If being an ethically sound company is at the top of the pyramid of your company's values, then granting an AI IRB the independence and power to veto proposals without the possibility of an appeal (to a member of your executive team, for instance) is a good idea.

Of course, that is often (sadly) not the case. Most companies see the AI IRB as a tool of risk mitigation, not elimination, and one should admit at least the possibility, if not the probability, of cases in which a company can pursue a project that is ethically risky while also highly profitable. For companies with that kind of ethical-risk appetite, either an appeals process will have to be created or, if they are only minorly concerned with ethical risk mitigation, they may make the pronouncements of the board advisory instead of required. At that point, though, they should not expect the board to be particularly effective at systematically mitigating ethical risks.

Values

You've assembled your AI IRB and defined its jurisdiction. Now you'll need to articulate the values by which it should be guided. The standard way of doing this is to create a set of principles and then seek to apply those principles to the case at hand. This is notoriously difficult, given the vast array of ways in which principles can be interpreted and applied; just think about the various and incompatible ways sincere politicians interpret and apply the principle of fairness.

In medical ethics—and in the law, for that matter—decision making is usually not guided by principles alone. Instead, the decision makers rely on case studies and precedent, comparing any given case under investigation with similar previous cases. This allows your IRB to leverage the insights brought to bear on the previous case to the present case and increases the probability of consistency in the application of principles across cases.

Progress can be made here by articulating previous decisions senior leadership made on ethical grounds before the existence of the IRB. Suppose, for instance, the IRB knows that senior leaders rejected a contract with a certain government due to ethical concerns about how the government operates or how they anticipated that government would use their product. The reasoning that led to the decision can reveal how future cases ought to be decided. In the event that no such cases exist and/or have been disclosed, it can be useful to entertain fictional examples, preferably ones that are not unlikely to become real examples in the future, and for the IRB to deliberate and decide on those cases. Doing that will ensure readiness for the real case when it arrives on their doorstep. It also encourages the cool objectivity with which fictional cases can be considered—when no money is on the line, for instance—to transfer to the real cases with which they'll be compared.

ADOPTING AN AI strategy is becoming a necessity to stay competitive. In a remarkable bit of good news, board members and data leaders see AI ethical risk mitigation as an essential component of that strategy. But current approaches are grossly inadequate, and many leaders don't know how to develop this part of their strategy.

In the absence of the ideal—a widespread commitment to creating a robust AI ethical risk program from day one—building, maintaining, and empowering an AI IRB can serve as a strong foundation for achieving that ideal. It can be created in relatively short order, piloted fairly easily, and built on and expanded to cover all product teams and even all departments, and creates and communicates a culture of ethics. That's a powerful punch not only for AI ethics but for the ethics of the organization more generally.

HBR Reprint H069JQ

Originally published on HBR.org April 1, 2021

Reid Blackman *is the founder and CEO of Virtue, an ethical risk consultancy. He works with companies to integrate ethical risk mitigation into the development, deployment, and procurement of emerging technology products. He also serves on numerous advisory boards and is a member of IEEE's Ethically Aligned Design Initiative. Prior to founding Virtue, he was a professor of philosophy at Colgate University and the University of North Carolina, Chapel Hill.*

Executive Summaries

"When companies make it clear that they are using AI to help people rather than to replace them, they significantly outperform companies that don't set that objective."

A BETTER WAY TO ONBOARD AI
PAGE 70

Competing with AI

Competing in the Age of AI

Marco Iansiti and Karim R. Lakhani
page 12

Today markets are being reshaped by a new kind of firm—one in which artificial intelligence runs the show. This cohort includes giants like Google, Facebook, and Alibaba, and growing businesses such as Wayfair and Ocado. Every time we use their services, the same thing happens: Rather than relying on processes run by employees, the value we get is delivered by algorithms. Software is at the core of the enterprise, and humans are moved off to the side.

This model frees firms from traditional operating constraints and enables them to compete in unprecedented ways. AI-driven processes can be scaled up very rapidly, allow for greater scope because they can be connected to many kinds of businesses, and offer very powerful opportunities for learning and improvement. And while the value of scale eventually begins to level off in traditional models, in AI-based ones, it never stops climbing. All of that allows AI-driven firms to quickly overtake traditional ones.

As AI models blur the lines between industries, strategies are relying less on specialized expertise and differentiation based on cost, quality, and branding, and more on business network position, unique data, and the deployment of sophisticated analytics.

HBR Reprint R2001C

The Business of Artificial Intelligence

Erik Brynjolfsson and Andrew McAfee
page 20

"General-purpose technologies," such as the internal combustion engine, have been the fundamental drivers of economic growth for 250 years. Artificial intelligence—particularly machine learning (ML)—is the most important such technology of our era. In the coming decade practically every industry will transform its key processes and business models to take advantage of ML. But not all expectations surrounding AI are realistic.

In this article, Erik Brynjolfsson and Andrew McAfee describe its real potential, its practical implications, and the barriers to its adoption. They note that AI skills are spreading quickly, through online educational resources as well as universities; the necessary algorithms and hardware for modern AI can be bought or rented as needed; and companies may not need all that much data to start making productive use of ML.

However, they say, ML is not without risks or limits: Machines may have hidden biases, derived from the data used to train them; neural networks deal with statistical rather than literal truths; and diagnosing and correcting system errors is often difficult, because a solution's underlying structure can be unimaginably complex.

HBR Article Reprint H03QXY
Big Idea Package Reprint BG1704

Adopting AI

Artificial Intelligence for the Real World

Thomas H. Davenport and Rajeev Ronanki | page 44

Cognitive technologies are increasingly being used to solve business problems; indeed, many executives believe that AI will substantially transform their companies within three years. But many of the most ambitious AI projects encounter setbacks or fail.

A survey of 250 executives familiar with their companies' use of cognitive technology and a study of 152 projects show that companies do better by taking an incremental rather than a transformative approach to developing and implementing AI, and by focusing on augmenting rather than replacing human capabilities.

Broadly speaking, AI can support three important business needs: automating business processes (typically back-office administrative and financial activities), gaining insight through data analysis, and engaging with customers and employees. To get the most out of AI, firms must understand which technologies perform what types of tasks, create a prioritized portfolio of projects based on business needs, and develop plans to scale up across the company.

HBR Reprint R1801H

Make Working from Home Work for Everyone

Whether you're new to remote work or you're a seasoned pro, you've likely experienced the benefits and the challenges of getting things done away from the office. The **You at Work: Doing Your Best Work Remotely** article collection provides best practices and tips on:

- Collaborating effectively with remote teams
- Maintaining focus while working from home
- Managing remote direct reports
- Running effective virtual meetings, and much more

YOU AT WORK: DOING YOUR BEST WORK REMOTELY
PRODUCT #BPWRKR

AVAILABLE EXCLUSIVELY AT:
store.hbr.org

Executive Summaries

Adopting AI

Getting AI to Scale

Tim Fountaine, Brian McCarthy, and Tamim Saleh | page 54

Most companies are struggling to realize artificial intelligence's potential to completely transform the way they do business. The problem is, they typically apply AI in a long list of discrete uses, an approach that doesn't produce consequential change. Yet trying to overhaul the whole organization with AI all at once is simply too complicated to be practical.

What's the solution? Using AI to reimagine one entire core business process, journey, or function end to end, say three McKinsey consultants. That allows each AI effort to build off the previous one by, say, reusing data or enhancing capabilities for a common set of stakeholders. An airline, for example, focused on its cargo function, and a telecom provider on its process for managing customer value.

Scaling up AI involves four steps: (1) Identify an area where AI will make a big difference reasonably quickly and there are multiple interconnected activities and opportunities to share technology. (2) Staff the team with the right people and remove the obstacles to their success. (3) Reimagine business as usual, working back from a key goal and then exploring in detail how to achieve it. (4) Support new AI-based processes with organizational changes, such as interdisciplinary collaboration and agile mindsets.

HBR Reprint R2103H

Inside Facebook's AI Workshop

Scott Berinato | page 62

Facebook has managed to harness and scale machine learning at a breakneck pace. One reason is that the company built an AI platform that allows any employee—deeply knowledgeable programmers and complete newbies alike—to take advantage of AI to achieve their goals.

In this article, HBR's Scott Berinato peeks behind the curtain of Facebook's Applied Machine Learning (AML) group, where Joaquin Candela oversees hundreds of employees running thousands of experiments a day. Candela's operating philosophy is that machine learning is not magic; rather it's a prosaic tool to be harnessed. Rather than recruiting the best talent to create the shiniest algorithms, Candela hones in on solving problems and creating value by enabling as many people as possible to take advantage of Facebook's huge trove of user data.

Candela describes the difficult early days of shifting the AML group's focus from seeking sophisticated algorithms to focusing on business impact by using more and better data and speeding up experimentation, and how he figured out where exactly AI fit on the product development path—key to building his AI factory. He details the intricacies of training AI to understand natural language or images, despite obstacles like verbal sarcasm or a picture of a cereal box (Is it food? Is it a box?) to create a product that can predict what a user will like or buy, supply appropriate search responses, and provide successful recommendations.

HBR Article Reprint H03QY6

Big Idea Package Reprint BG1704

A Better Way to Onboard AI

Boris Babic, Daniel L. Chen, Theodoros Evgeniou, and Anne-Laure Fayard
page 70

In a 2018 Workforce Institute survey of 3,000 managers across eight industrialized nations, the majority of respondents described artificial intelligence as a valuable productivity tool. But respondents to that survey also expressed fears that AI would take their jobs. They are not alone. The *Guardian* recently reported that in the UK "more than 6 million workers fear being replaced by machines."

AI's advantages can be cast in a dark light: Why would humans be needed when machines can do a better job? To allay such fears, employers must set AI up to succeed rather than to fail. The authors draw on their own and others' research and consulting on AI and information systems implementation, along with organizational studies of innovation and work practices, to present a four-phase approach to implementing AI. It allows organizations to cultivate people's trust—a key condition for adoption—and to work toward a distributed cognitive system in which humans and artificial intelligence both continually improve.

HBR Reprint R2004C

The Future of Jobs

Collaborative Intelligence: Humans and AI Are Joining Forces

H. James Wilson and Paul R. Daugherty
page 84

Artificial intelligence is transforming all sectors of the economy, but there's no reason to fear that robots will replace all human employees. In fact, companies that automate their operations mainly to cut their workforces will see only short-term productivity gains, say the authors. Their research, involving 1,500 firms in a range of industries, shows that the biggest performance improvements come when humans and smart machines work together, enhancing each other's strengths.

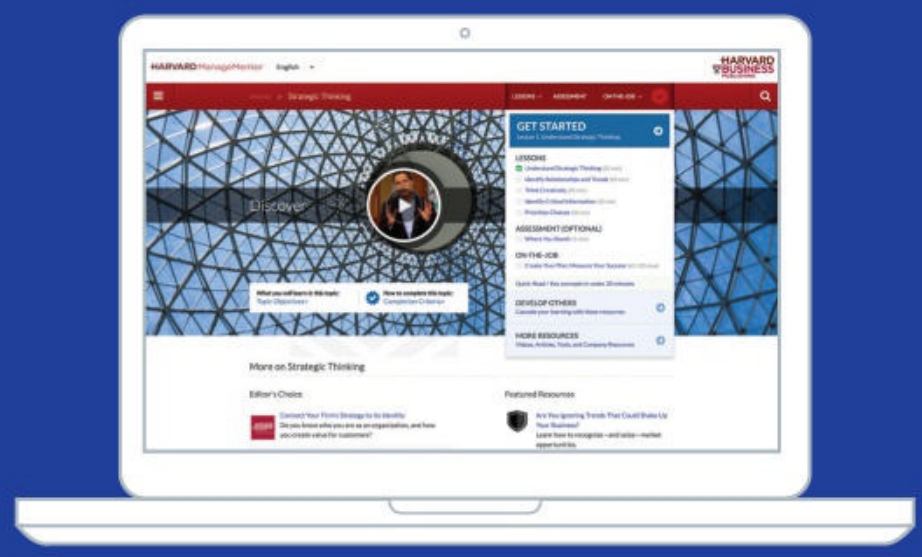
People need to train AI agents, explain their outputs, and make sure they are used responsibly. AI agents, in turn, can assist people with information gathering, data crunching, routine customer service, and physical labor, thereby freeing them for higher-level tasks that require leadership, creative thinking, judgment, and other human skills.

To get the most out of AI, companies need to redesign their business processes. After deciding what needs improvement—their operational flexibility, speed, or scalability; their decision-making; or their ability to personalize products and services—they can devise appropriate solutions. That will mean not only implementing AI technology but also developing employees who can work effectively at the human-machine interface.

The authors describe how a number of firms are already taking these steps and optimizing collaborative intelligence. But many more should follow their example.

HBR Reprint R1804J

Harvard Business Review



Everything You Need to
Succeed as a Manager

Harvard ManageMentor gives you the skills and solutions you need to effectively meet management challenges every day. Our newest version of this premier online learning tool is packed with relevant, continuously updated content that will let you gain invaluable skills for yourself, your team, and your organization—at the pace you want, when you want it.

HARVARD MANAGEMENTOR ONLINE TOOL | 41 TOPICS | #5678

hbr.org/harvardmanagementor
1-800-988-0886 OR +1-617-783-7500

Executive Summaries

Working with Machines

Learning to Work with Intelligent Machines

Matt Beane | page 112

Although companies spend billions on formal training for employees, most of the skills needed to perform a specific job can be learned only by doing it. This on-the-job learning (OJL) has long depended on mentorship, with experts coaching apprentices. But today OJL is under threat from the headlong introduction of sophisticated analytics, AI, and robotics into many aspects of work. These technologies are moving trainees away from learning opportunities and experts away from the action. The author describes the "deviant," rule-breaking work-arounds— "shadow learning"—that surgeons in training, police officers, M&A analysts, and others are figuring out on their own to overcome these obstacles and suggests how companies can benefit from studying them.

HBR Reprint R1905K

Algorithms Need Managers, Too

Michael Luca, Jon Kleinberg, and Sendhil Mullainathan | page 120

Algorithms are powerful predictive tools, but they can run amok when not applied properly. Consider what often happens with social media sites. Today many use algorithms to decide which ads and links to show users. But when these algorithms focus too narrowly on maximizing click-throughs, sites quickly become choked with low-quality content. While clicks rise, customer satisfaction plummets.

The glitches, say the authors, are not in the algorithms but in the way we interact with them. Managers need to recognize their two major limitations: First, they're completely literal; algorithms do exactly what they're told and disregard every other consideration. While a human would have understood that the sites' designers wanted to maximize quality as measured by clicks, the algorithms maximized clicks at the expense of quality. Second, algorithms are black boxes. Though they can predict the future with great accuracy, they won't say what will cause an event or why. They'll tell you which magazine articles are likely to be shared on Twitter without explaining what motivates people to tweet about them, for instance.

To avoid missteps, you need to be explicit about all your goals—hard and soft—when formulating your algorithms. You also must consider the long-term implications of the data the algorithms incorporate to make sure they're not focusing nearsightedly on short-term outcomes. And choose the right data inputs, being sure to gather a wide breadth of information from a diversity of sources.

HBR Reprint R1601H

Keeping AI from Going Wrong

When Machine Learning Goes Off the Rails

Boris Babic, I. Glenn Cohen, Theodoros Evgeniou, and Sara Gerke | page 132

Products and services that rely on machine learning—computer programs that constantly absorb new data and adapt their decisions in response—don't always make ethical or accurate choices. Sometimes they cause investment losses, for instance, or biased hiring or car accidents. And as such offerings proliferate across markets, the companies creating them face major new risks. Executives need to understand and mitigate the technology's potential downside.

Machine learning can go wrong in a number of ways. Because the systems make decisions based on probabilities, *some* errors are always possible. Their environments may evolve in unanticipated ways, creating disconnects between the data they were trained with and the data they're currently fed. And their complexity can make it hard to determine whether or why they made a mistake.

A key question executives must answer is whether it's better to allow smart offerings to continuously evolve or to "lock" their algorithms and periodically update them. In addition, every offering will need to be appropriately tested before and after rollout and regularly monitored to make sure it's performing as intended.

HBR Reprint R2101F

When you offer a modern approach, you just seem to stand out.

Schwab offers modern wealth management solutions that combine technology with human help when you need it, in an approach that's different than most. With Schwab, you get transparent pricing and innovative investing platforms backed by 24/7 live customer support and a Satisfaction Guarantee, at the lowest costs in the industry. It's time to change to a modern approach to wealth management.

charles SCHWAB

Investor's Business Daily MOST TRUSTED FINANCIAL COMPANIES 2020 #1 Overall

Own your tomorrow.

Learn more at **Schwab.com**

If you are not completely satisfied for any reason, at your request Charles Schwab & Co., Inc. ("Schwab") or Charles Schwab Bank, SSB ("Schwab Bank"), as applicable, will refund any eligible fee related to your concern. Refund requests must be received within 90 days of the date the fee was charged. Schwab reserves the right to change or terminate the guarantee at any time. Go to schwab.com/satisfaction to learn what's included and how it works.

From Investor's Business Daily, October 12, 2020, ©2020 Investor's Business Daily, Inc. All rights reserved.

©2021 Charles Schwab & Co., Inc. All rights reserved. Member SIPC. (ADP112778-00) 0521-1VG8

WHAT'S EASY TO SEE IN OUR CLOUD? OPPORTUNITY.

It's easier than ever to access advanced cloud analytics with SAS® Viya® on Microsoft Azure. Viya is optimized for Azure, which means it's easier to find, share and analyze information all in one place. So more people can ask the questions that lead to more opportunities. Together, SAS and Microsoft make it easier to find answers in the cloud.

Learn how at sas.com/microsoft

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries.
® indicates USA registration. Other brand and product names are trademarks of their respective companies. © 2021 SAS Institute Inc. All rights reserved. G191288.1021