Expert Insights

–

Proven concepts for scaling AI

From experimentation to engineering discipline

IBM **Institute for Business Value**

Experts on this topic

Beth Rudden

Distinguished Engineer
Principal Data Scientist,
Cognitive and AI, IBM Services
linkedin.com/in/brudden/
brudden@us.ibm.com

Beth Rudden transforms people and companies through analytics and empowerment with the ethical use of data. She leads large, geographically dispersed teams to develop cognitive analytic solutions that deliver actionable insights for IBM's clients. Beth has received patents on solutions that provide clearer insights, better customer understanding, and faster implementation. Her background in anthropology, language, and data science also helps her develop models for the transformation of the IBM services workforce.

Wouter Oosterbosch

Chief Data Scientist–Europe
EU Leader, Worldwide Advanced Analytics Center of Competence, IBM Services
linkedin.com/in/wouteroosterbosch/
w.oosterbosch@nl.ibm.com

Wouter Oosterbosch is a neuroscientist by training, which has cultivated a keen interest in wherever humans and data intersect. He is an experienced cross-industry data science leader who has seen companies through the various phases of AI implementations and empowered teams across the globe: from a disorganized and undocumented data jungle to delivering actionable, scalable, and trusted results for clients.

Dr. Eva-Marie Muller-Stuler

Chief Data Scientist—Middle East/ Africa, Advanced Analytics and AI Practice Leader, IBM Services
linkedin.com/in/dr-eva-marie-muller-stuler-02ab5946/
Eva-Marie.Muller-Stuler@ibm.com

Dr. Muller-Stuler has more than 15 years of experience in leading large-scale business transformations as well as numerous data science and AI projects globally. She has pioneered highly successful solutions with governments and top-tier organizations alike.
Dr. Muller-Stuller is also a trusted advisor to governments on change management through AI.

Key takeaways

AI growth continues apace

The number of AI adopters has grown 65 percent in the past four years, a trend the pandemic's business disruption is accelerating relative to other technology priorities.[1]

AI demands engineering discipline

Companies need to embrace AI holistically to address the scaling problem—rooting it in business strategy, innovation, and competitive differentiation, then deeply integrating it into evolving business operating models and workflows.

AI proofs-of-concept (POCs) must evolve

As AI technology continues to mature, many concepts have been proven, so organizations can redirect early-stage trials toward market-ready pilots.

—

The proof-of-concept (POC) is dead; long live the POC!

Poor, misunderstood artificial intelligence (AI). It's alternately overhyped as a digital nirvana or vilified as a dystopian menace. Yet in the pragmatic here and now, it is neither.

Primarily, AI is a way to augment human capabilities and performance, creating better outcomes for people—customers, employees, partners, and other stakeholders—and better financial returns for businesses. Think human aid, not humanoid.

Some organizations view AI as a means to achieving incremental but tangible outcomes with intelligent workflows—more efficient business operations, more compelling customer experiences, and more insightful decision-making—so human ingenuity and empathy can take center stage. Others have embraced the more transformational nature of AI, which has resulted in new business models, novel approaches to addressing business disruptions (such as the COVID-19 pandemic), and radical improvements in workflow performance.

The growing adoption of AI is reflected in metrics the IBM Institute for Business Value (IBV) has tracked biannually since 2016. Data from thousands of C-level business executives across regions, industries, and functions points to a trend we expect to accelerate modestly as a result of the pandemic:

- Companies with AI in active use have *increased* from 26 percent four years ago to 44 percent in 2020 (a more conservative view than some estimates).[2]
- In the midst of the pandemic, 84 percent of total organizations expect a *similar or higher* level of organizational focus on AI.[3]
- Nearly one-third plan to *boost* their investment in AI as a result of the pandemic.[4]

These trends are consistent with other recent estimates, with IDC forecasting that worldwide spending on AI will increase in 2020–in contrast to a decline in overall IT spending–and double in the next four years.[5]

> “AI and ML are only beginning to emerge from their formative stage–and the peak of the hype cycle–into a period of more practical, efficient development and operations.”[6]

VC firm Andreessen Horowitz

But successful scaling—shepherding AI projects from sandboxes to pilots and minimum viable products (MVPs) all the way to industrial-strength commercialization in the business—has bedeviled many companies. As the IBV observed in mid-2018, “organizations are knee-deep in AI pilots and proofs-of-concept...and foraying piecemeal into exciting but isolated use cases”—a reality later acknowledged by many other market observers.[7]

Yet even now, 90 percent of companies have difficulty scaling AI across their enterprises. So, it’s not surprising that about half of AI projects fail.[8]

To be sure, AI is a complex, multi-faceted business and technological innovation with layers of interconnected and moving parts. No one aspect can single-handedly ensure success in moving AI projects into commercial use. There is no silver bullet, no panacea.

Plain vanilla “change management” won’t cut it. Neither will an anodyne “alignment with business strategy.” Not even tried-and-true “process improvement” or more recent “agile methodologies” will suffice—no matter how many sigmas and spaghetti charts or scrums and sprints are marshalled.

What’s needed is a step change in the role of AI: a shift from being viewed at arm’s length as the latest incarnation of technological wizardry to a strategic capability embedded throughout the business. From proof-of-concept to proof point.

Companies need to stop chasing data science experiments pell-mell and start embracing AI thoughtfully and holistically—rooting it in business strategy, innovation, and competitive differentiation, then deeply integrating it into evolving business operating models and workflows, organizational structures and governance, data architecture and infrastructure, and even their cultural values and ethics.

To advance, organizations first must treat AI as a discipline—with robust engineering and ethical principles, rigorous operations and governance, and an adaptable approach that emphasizes pragmatism over theory. There are readily available tools to help achieve this. Organizations also must apply greater focus on scientific innovation—with R&D-like capabilities that continuously explore the bleeding edge in order to differentiate.

Of course, progress is rarely linear. There will be projects that succeed in early stages, but still fail to achieve human adoption. There is value gained from AI pilots and MVPs: perfect is still sometimes the enemy of good enough. However, beta tests need to be developed and launched as part of a commercialization engine explicitly designed and geared toward growth and scale.

Otherwise, companies risk becoming mired in endless cycles of experimentation, ever dabbling but never doing.

Getting serious with AI engineering and operations

For businesses in earlier stages of AI adoption, the need to treat AI as a discipline may not be obvious. However, it demands the same communication, structure, and rigor common in more established areas of a business to realize value fully.

All too often, model development takes place on a data scientist’s laptop, and orchestration is done manually, or ad hoc, using custom code and scripts. This is much the same way traditional application development took place before agile DevOps best practices emerged.

The net effect is that data teams—scientists, engineers, and others—often are forced to work inefficiently. They are burdened with manual tasks, such as handoffs to developers within whose apps their machine learning (ML) models will ultimately run. This creates an impediment to using ML models on the same cadence—and using the same DevOps processes—as applications. It slows the delivery of ML-enabled applications and reduces the business returns on AI investments.

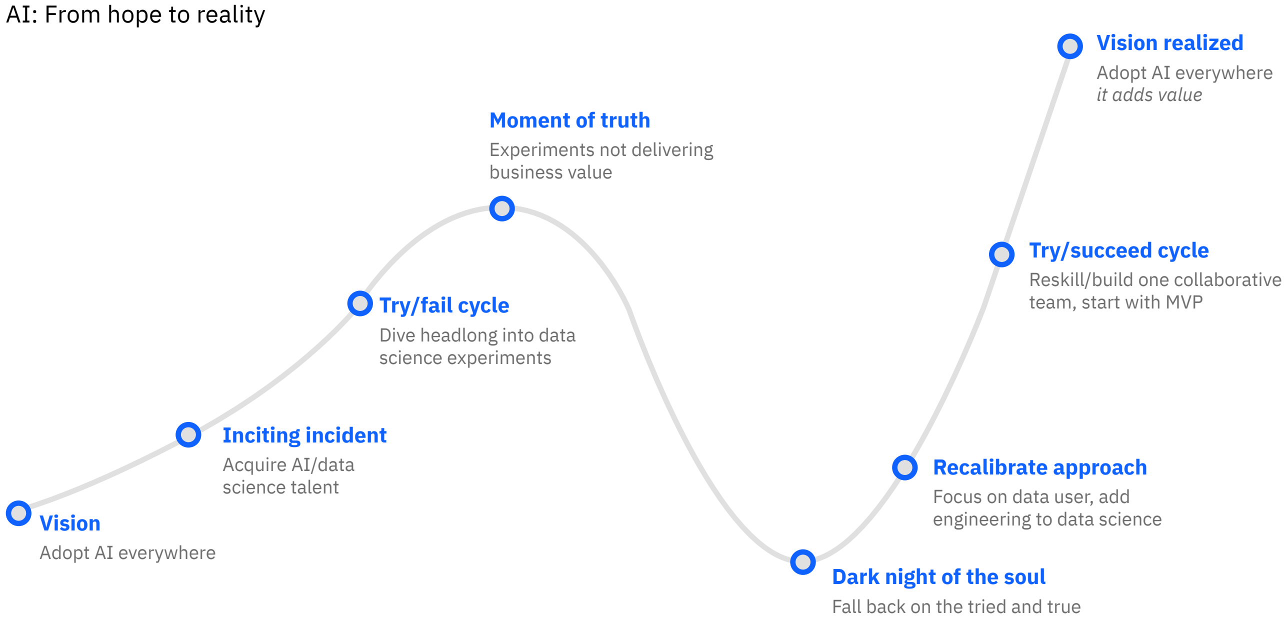
Another reason AI initiatives stall before going into production is that the projects are often siloed, with a wall or gap between developers and stakeholders. This is exacerbated when it is unclear who owns and who controls specific data. Moreover, some AI teams are relatively new, with roles and responsibilities still uncertain, various "tribal" allegiances, and even disparate tool sets in the same organization.

Even more established teams need to interface with diverse constituencies and stakeholders. All this can make clear, precise communication challenging.

The track of a company's all-too-typical AI program is captured, perhaps with a hint of cynicism born from years of experience and observation, in Figure 1. We believe the trough can be largely avoided, though, with a more structured approach.

Figure 1

AI: From hope to reality



Source: IBM analysis

Agile DevOps + automated ITOps + MLOps = AIOps

We call the approach to establishing methodical rigor "AI engineering and operations"—with four high-level focus areas, as well as underlying principles, processes, and tools to guide AI initiatives to production at scale (see Figure 2).

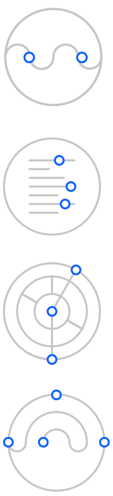
Even for companies with advanced data science and analytics capabilities, as well as enlightened approaches to software engineering, AI engineering and operations may necessitate the creation of new roles (such as ML engineers and AI operational specialists) as part of merging the different types of development cycles. And a more nuanced approach to solution design that features dynamic feedback loops—straddling the development and production environments —may challenge the "comfort zones" of traditional architects more accustomed to robustness at all costs.

Similar to how many companies use DevOps and other software engineering approaches, AI engineering and operations extends the proven benefits of decreased development cycles, improved collaborations, higher levels of operational efficiency, and more effective deployment (see sidebar, "Red Hat: AI in software with open source concepts.") The approach creates an environment that brings a structured focus to ushering projects through development into production and, ultimately, delivering commercial results.

—

Figure 2

AI engineering and operations



Design
A human-AI experience designed for usability, as well as standard toolsets and methodologies to enhance speed-to-value and quality standards across AI projects.

Deployment
A framework that automates deployment to improve efficiency and auditability.

Monitoring
Technical and quality key performance indicators (KPIs) and processes to regularly measure and benchmark them.

Embedding
Methods to check for bias in models, tools that visualize decisions taken by AI models, and broader company ethical guidelines.

Source: IBM analysis

Red Hat: AI in software with open source concepts

Like any typical technology company, Red Hat showed an early interest in AI and ML, exploring how it could use these technologies to apply to its products and services to benefit customers.

That all changed about four years ago. That's when Red Hat started to intensify its focus on AI as part of a broader portfolio approach to its offerings to make sure they could interoperate and support customers' growing demand for AI and ML workloads on containers and Kubernetes.

Red Hat increasingly operationalized AI on top of its platforms, establishing what would become the foundation for Open Data Hub, a meta-project grounded in AI engineering principles that integrates open source projects into practical solutions complemented by AI ecosystem partners. The open source community can experiment and develop intelligent applications without incurring high costs and mastering the complexity of modern ML and AI software stacks.

To give structure to and help deliver on its strategy, Red Hat formed the AI Center of Excellence (CoE). This organization was augmented with a newly formed "forward deployed engineering team," mobilizing its top data scientists to provide innovation and value to customers through a services engagement model. As this effort grew, engineering discipline was added. DevOps and agile methods were used to strengthen and formalize the company's approach to AI development.

Red Hat now uses "Open Innovation Labs" to collaborate with customers on AI/ML projects that employ best-of-breed open source technologies.[9] For example:

- For an automotive client, the AI CoE helped develop a platform to deliver faster, more accurate driving simulations and data analytics with scalable ML and big data processing capabilities. The platform was configured and created in just three months.[10]
- For a healthcare client, the AI CoE created a prediction and therapy optimization platform to collect and analyze clinical data and signal caregivers in real time to initiate early care.[11]

For Red Hat, open source has taken on a new life in the context of a structured approach to AI.

Innovating with NLP and semantics

Many companies are integrating ML and deep learning into their business operations, but the resulting models and algorithms are too often based solely on structured data. And one of the top challenges cited by companies in leveraging information is knowledge locked in unstructured data.

There is a better choice. Using advanced natural language processing (NLP) capabilities to enrich AI models with unstructured data can be challenging but helps provide the human context for how data is viewed and used by people. In other words, it brings the power of human language to bear. And with a deeper integration of these capabilities, AI now has two symbiotic learning loops instead of just one: a semantically-enabled learning loop for data integration and a statistically-enabled learning loop for ML.

At the 2020 US Open Tennis Championships, NLP and semantic technologies powered a popular improvement to the fan experience in an otherwise remote environment (see sidebar, “US Open: Open questions? Case closed”).

Semantic technologies and NLP also provide the necessary lineage and provenance that allow developers to verify that the AI system understands what people are writing or saying. And using the proper tools and algorithms for each situation is at the heart of creating more intelligent workflows.[12]

US Open: Open questions? Case closed

At this year’s US Open Tennis Championships, like at so many other sports events, spectators couldn’t fill the seats as part of the COVID-19 protocols—a decision made by the United States Tennis Association (USTA) in mid-June. But thanks to AI and other leading technologies, there were new ways for fans to be a part of the tennis experience.[13]

For example, an online application called “Open Questions” allowed fans to contribute their arguments to the debate about various tennis topics. With topics such as the best player of all time and the greatest rivalry, the application used NLP to analyze millions of sources and delivered a debate-like pro/con argument. Fans shared their opinions, adding to the debate.[14]

Each day during the US Open, which occurred in the late summer of 2020 in New York City, fans could add their opinions to a topic, and that input was added to the database. New computer-based, AI-generated narrations around each topic were produced daily from the updated data, creating continuously deeper, more meaningful debates.

Another project was an AI-powered “cheat sheet” available to fans for every match. “Match Insights” used NLP technology to analyze millions of unstructured data sources like articles, blogs, and expert opinions. It pulled key insights from that mountain of data and converted them into a brief narrative form, enabling fans to get information ahead of matches.[15] And this comparative analysis was presented in natural language, serving statisticians and casual fans alike.

Fans hope to fill the stands next year to enjoy tennis mastery and cheer for their favorite players. Until they can safely return, and probably even after that, NLP will enhance their tennis experience.

Making the case for building AI capabilities

For organizations that want to capture the true potential of AI in production, the first step in a more thoughtful and holistic approach is to put the discipline of AI engineering and operations at the core. And for those companies ready to take the next step in innovating with AI, building a robust NLP/semantic capability is the next step in achieving greater human understanding through AI.

Moreover, not embracing this approach carries the risk of allowing the gap between data scientists and operations teams to widen. And the further down the road ML projects go without being governed by sound operational procedures, the less likely they will be to make the grade.

As such, we cannot lose sight of the people who will make AI, use it, and benefit from it. We need bold *leaders* so the AI vision for the opportunity is clear and ethical (see sidebar, "Insights: Ethics in AI"). We need inspired *designers* so our human relationship with AI and its conditions will thrive. We need thoughtful *engineers* so the output engenders confidence and trust.

And we need to keep in mind that the *humans* who interact with AI are perhaps the most important part of any project team. Ultimately, they are responsible for realizing the actual, not just ideal, individual experiences, intelligent workflows, collaborative decisions, and tangible business value.

—

Figure 3

Seven key requirements for trustworthy AI

- Human agency and oversight
- Technical robustness and safety
- Privacy and data governance
- Transparency
- Diversity, non-discrimination, and fairness
- Societal and environmental well-being
- Accountability

Source: European Commission, High-Level Expert Group on AI, "Ethics Guidelines for Trustworthy Artificial Intelligence."

Insight: Ethics in AI

Critical areas of judgment—especially decisions that directly impact others' lives and well-being—are governed by standards of appropriate action. But the ethical parameters around AI remain undefined and vague, in some instances pushed aside as impediments to progress.

In a study based on a survey of 1,250 C-level executives, the IBM Institute for Business Value found that more than half of the executives surveyed say AI actually could improve their companies' ethical decisions. A majority also say AI could be harnessed as a force for societal good, not just for good business. And nearly all the respondents currently adopting AI are formally considering ethics as part of their AI initiatives.[16] But first, the right ethical framework has to be in place.

While most major technology firms have issued their own guidelines, some have explicitly endorsed those from the European Commission's High-Level Expert Group. These guidelines define a human-centric "trustworthy" AI approach built around seven requirements (see Figure 3).

Our survey findings also suggest the need for more corporate education about and engagement with AI ethics issues. The World Economic Forum's AI Board Toolkit, developed through collaboration with various public and private partners including IBM, is a start in this direction.

As is appropriate for AI, not all the burden need be on the people involved. There are effective tools and infrastructure that can be deployed to continually monitor AI systems for trustworthiness and to avoid potential ethical lapses.

Yet corporate education, professional standards, and effective tools are not enough. There are significant questions about the tradeoffs between individual privacy and business value, regulation and innovation, and transparency and competitive advantage. Those tradeoffs deserve to be debated in a thoughtful and collaborative manner.

What's at stake may be no less crucial than a wholesale rethinking of the social contract.

Action guide

AI engineering and operations

While getting AI out of the lab and into full production is far from a trivial undertaking, we have identified key actions businesses can take to speed the path to scaling AI.

First, here are leading practices for less established AI adopters (companies in the considering, evaluating, and piloting phases of AI):

Get started

Development can often take place in "bite-sized chunks" in parallel. At the same time, understanding what data you have, where it resides, and who manages it can go a long way to increasing confidence in the outputs. AI does not necessarily require an initial massive data governance project to curate and cleanse data.

Start small but design for scale

Use an MVP to lay the foundation for something larger. Initial projects should be prioritized based on business impact, complexity, and risk. From there, you can build scale. Create and follow a roadmap based on impact and feasibility. If a pilot doesn't succeed, accept it as a learning process and move on. Don't expect every project to move into full production. A hybrid multicloud environment lends itself to scaling with data from various sources.

Adopt engineering principles

If you already are using DevOps or other software engineering approaches, establish a small team to transfer those skills and processes to AI projects. Adjust these policies and processes for the nuances of an AI environment.

Establish measurements for success

If it's worth doing, it's worth measuring. Metrics should be mapped against key success factors and significant risks. They also should be open and transparent, allowing the relevant internal teams to review progress. Feedback loops should provide input for new design and development. In AI, failure *is* an option, as long as companies learn from those constructive failures.

Appoint strong leadership

Confirm all AI projects support the strategic agenda and are designed with customers and other stakeholders in mind. AI should be regularly tested for bias and transparency to help ensure the output is ethical and fair. Leaders should also be responsible for building or acquiring the requisite AI skills and training.

Next, here are leading practices for more established AI adopters (companies in the implementing, operating, and optimizing phases of AI):

Establish an AI playbook

The playbook should be a living document, with checklists and engineering principles, built upon successes, failures, and KPIs. Create an architecture and team structure that operates at the intersection of design and data centers.

Continuously document and improve

Reinforce that deploying AI models is not the only goal or the end of a project. For AI to scale, you need to continuously evaluate and improve your models while they are in production. If it's not repeatable, it's not reliable—and documentation is critical for repeatability.

Monitor models

On an ongoing basis, monitor the explainability, fairness, and robustness of your AI models. Develop inspection algorithms—ethical "bots"—that serve as virtual microscopes to search for unintended bias and other issues.

Innovate at scale

Adopt and integrate deep, robust NLP capabilities and other forward-looking elements of AI matched to distinct use cases that add clear business value. Integrate disparate internal and external data sources. Adopt the mindset of an AI startup. Consider assigning some resources to explore bleeding edge technologies.

Engage ecosystem partners

Consider partnering with others to establish and/or influence relevant standards, drive transparency, and foster trust. Engage academics, think tanks, startups, and other trusted third parties.

Notes and sources

1 Unpublished data from IBM Institute for Value survey on AI Value (n=6700 C-level business executives in major industries, functional areas, and geographies).

2 Ibid; Ammanath, Breena, David Jarvis, and Susan Hupfer. "Thriving in the era of pervasive AI." Deloitte. 2020. https://www2.deloitte.com/xe/en/insights/focus/cognitive-technologies/state-of-ai-and-intelligent-automation-in-business-survey2.html

3 Unpublished data from IBM Institute for Value survey on AI Value (n=2765 C-level business executives in major industries, functional areas, and geographies).

4 Ibid.

5 "Worldwide Spending on Artificial Intelligence Is Expected to Double in Four Years, Reaching $110 Billion in 2024, According to New IDC Spending Guide." IDC. September 25, 2020. https://www.idc.com/getdoc.jsp?containerId=prUS46794720

6 Casado, Martin and Matt Bornstein. "Taming the Tail: Adventures in Improving AI Economics." Andreessen Horowitz. August 12, 2020. https://a16z.com/2020/08/12/taming-the-tail-adventures-in-improving-ai-economics

7 Christopher, Elena, Glenn Finch, Brian C. Goehring, Cathy Reese, Thomas Reuner, and Yashih Wu. "Artificial intelligence: The killer app for data." HFS Research and IBM Institute for Business Value. July 2018 / February 2019. https://www.ibm.com/thought-leadership/institute-business-value/report/killerappdata#; Brenna, Francesco, Giorgio Danesi, Glenn Finch, Brian C. Goehring, and Manish Goyal. "Shifting toward Enterprise-grade AI: Confronting skills and data challenges to realize value." IBM Institute for Business Value. September 2018. https://www.ibm.com/thought-leadership/institute-business-value/report/enterpriseai; Awalegaonkar, Ketan, Robert Berkey, Greg Douglass, and Athena Reilly. "AI: BUILT TO SCALE." Accenture. November 14, 2019. https://www.accenture.com/us-en/insights/artificial-intelligence/ai-investments; Justice, Cliff, Todd Lohr, Martin Sokalsi, Vinodh Swaminathan, Matt Fish, Brad Fisher, and Traci Gusher. "AI transforming the enterprise: 8 key AI adoption trends." KPMG. 2019. https://advisory.kpmg.us/content/dam/advisory/en/pdfs/2019/8-ai-trends-transforming-the-enterprise.pdf

8 Linthwaite, Rachel." Overcome Obstacles To Get To AI At Scale." Forrester. January 2020. https://www.ibm.com/downloads/cas/VBMPEQLN; "IDC Survey Finds Artificial Intelligence to be a Priority for Organizations But Few Have Implemented an Enterprise-Wide Strategy." IDC. July 8, 2019. https://www.idc.com/getdoc.jsp?containerId=prUS45344519

9 "Red Hat Services Solution: Open AI/ML Platform." Red Hat Datasheet. 2020. https://www.redhat.com/cms/managed-files/co-services-open-ai-platform-datasheet-f21162pr-202002-en.pdf

10 "Global automotive group races to automated driving with data platform." Red Hat. Accessed August 30, 2020. https://www.redhat.com/en/success-stories/bmwgroup

11 "HCA Healthcare uses innovative data platform to save lives." Red Hat. Accessed August 30, 2020. https://www.redhat.com/en/success-stories/hca-healthcare

12 "Building the Cognitive Enterprise: Nine Action Areas." IBM Institute for Business Value. May 2020. https://www.ibm.com/thought-leadership/institute-business-value/report/build-cognitive-enterprise#

13 Baughman, Aaron, Gray Cannon, Micah Forster, and Nick Wilkin. "At the US Open: Machine writing and discovery with Watson." IBM. August 28, 2020. https://developer.ibm.com/components/watson-discovery/articles/at-the-us-open-machine-writing-and-discovery-with-watson/

14 "USTA to try to Bring Fans to US Open Virtually along with the US Open Experience into Fans Homes." Tennis Panorama. August 28, 2020. https://www.tennispanorama.com/archives/70866

15 Ibid.

16 Goehring, Brian, Francesca Rossi, and David Zaharchuk. "Advancing AI ethics beyond compliance: From principles to practice." IBM Institute for Business Value, April 2020. https://www.ibm.com/thought-leadership/institute-business-value/report/ai-ethics#

About Expert Insights

Expert Insights represent the opinions of thought leaders on newsworthy business and related technology topics. They are based upon conversations with leading subject matter experts from around the globe. For more information, contact the IBM Institute for Business Value at iibv@us.ibm.com.

IBM Corporation
New Orchard Road
Armonk, NY 10504
Produced in the United States of America
September 2020

93035193USEN-01