playbook

Next Generation Data Science: The AI-Driven Data Science Life Cycle

By Donald Farmer and Fern Halper, Ph.D.

playbook

Table of Contents

The Rise of AI and the Emergence of Open-Source Tools

THE STRATEGIC ROLE OF AI AND DATA SCIENCE

Artificial intelligence is no longer an experimental technology, but a strategic imperative for almost every organization, reshaping how they operate, compete, and create value. For enterprises, AI enables unprecedented personalization of customer experiences, transforms decision-making through predictive analytics, and drives operational efficiency across every business function. From supply chain optimization to fraud detection, applications powered by AI are generating measurable competitive advantages for early adopters.

Although the excitement is centered on AI itself, data science is still a critical enabler as the discipline that transforms raw information into actionable intelligence. However, effective data science demands significant expertise spanning statistics, programming, domain knowledge, and business acumen. This multidisciplinary requirement creates both opportunity and challenge for IT leaders.

TDWI research indicates a bifurcated landscape. Approximately half of the organizations we survey are actively developing AI models and establishing mature data science practices, while the remainder are working toward these capabilities but facing implementation barriers. This divide largely stems from the skills gap challenge—the shortage of professionals who can bridge technical complexity with business value creation.

The technical expertise barrier is particularly acute as data science requires proficiency in programming languages, statistical methods, machine learning algorithms, and data engineering practices. Organizations struggle to recruit and retain teams with these skills, limiting their capacity to adopt AI. So the challenge is not purely a technical problem, but an organizational one.

THE OPEN SOURCE ADVANTAGE AND BARRIER

Open source tools have become the backbone of modern data science, with Python and R dominating the landscape due to their flexibility, cost-efficiency, and rich ecosystems. These platforms offer access to cutting-edge algorithms and methodologies, often months or years ahead of commercial alternatives.

> *Open source tools provide a low-cost entry point for advanced analytics, with freely available code supported by a vibrant community.*

Open source offers a range of compelling benefits that make it attractive to organizations and developers alike. It provides a low-cost entry point for advanced analytics, with freely available code supported by a vibrant community. The foundational libraries powering today's AI revolution (scikit-learn for machine learning, TensorFlow and PyTorch for deep learning, and R's tidyverse for data manipulation) personify open source's innovation velocity. Additionally, open source eliminates vendor lock-in, giving users the flexibility and autonomy to manage their tools and updates without being tied to proprietary software providers.

However, this advantage comes with challenges. Open-source tools often require technical expertise. Unlike commercial platforms with integrated support structures, open source solutions demand in-house knowledge for troubleshooting, maintenance, and optimization. Additionally, enterprise governance requirements (audit trails, security controls, compliance documentation) often require custom implementation in open-source environments.

The tension between open source's potential for innovation and enterprise operational needs creates a strategic dilemma for IT leaders—embrace the complexity for maximum flexibility or accept limitations for operational simplicity?

HOW AI TOOLS CAN HELP

Emerging AI-infused data science platforms promise to resolve this dilemma by embedding intelligent assistance directly into open source workflows. Some of these capabilities are found in commercial tools now. These solutions can help maintain the flexibility and cost advantages of Python and R while reducing technical barriers and accelerating productivity. AI-infused tools can help across the data science life cycle in numerous areas, including data preparation, feature engineering, code generation and completion, insight generation, workflow integration, and monitoring and maintenance.

For instance, AI-powered tools can automatically profile data sets, identify quality issues, detect anomalies, and suggest appropriate transformations. They can leverage machine learning to recommend relevant variables and transformations for feature engineering. They can provide real-time assistance for script writing, debugging, and optimization. They can provide drift detection, performance monitoring, and alerting systems to ensure models remain effective over time without constant manual oversight.

The ultimate promise behind these capabilities is to democratize data science capabilities while maintaining professional rigor and enabling broader organizational participation in AI initiatives without compromising quality or governance standards. For IT leaders, this represents an opportunity to accelerate AI adoption while leveraging existing open source investments and addressing skills constraints through intelligent automation.

The Business Case for AI-Infused Data Science

The convergence of artificial intelligence with traditional data science workflows represents more than technological evolution; it signals a fundamental shift in how organizations can scale analytical capabilities while managing complexity and risk. AI-infused tools don't replace human expertise; instead, they amplify it by automating routine tasks, suggesting optimizations, and providing intelligent guidance throughout the analytical process. These platforms maintain the flexibility and innovation advantages

of open source environments while introducing intelligent automation that reduces technical barriers, accelerates development cycles, and improves output quality.

For IT leaders, this represents an opportunity to democratize advanced analytics capabilities across the organization while maintaining governance standards and leveraging existing technology investments. The business case can extend beyond cost savings to include strategic agility, risk mitigation, and competitive advantage through faster time-to-insight and broader organizational participation in data-driven decision-making.

FRAMING THE DATA SCIENCE LIFE CYCLE

Understanding the business value of AI-infused tools requires recognizing the complexity and interdependencies within the data science life cycle. This life cycle represents a systematic approach to transforming business questions into actionable insights through data analysis and predictive modeling. There are four main components to this life cycle: model preparation, model building, model operationalization, and model governance. Each phase presents distinct challenges, skill requirements, and opportunities for intelligent automation.

Model Preparation

This foundational phase involves problem definition, data identification, and initial preparation activities. First, problem definition and business alignment are the process of translating business objectives into analytically tractable questions, which requires both domain expertise and technical understanding. Organizations frequently struggle here, with projects failing due to misaligned expectations or poorly scoped analytical objectives.

Data identification, acquisition, and initial assessment follow, involving the discovery and evaluation of relevant data sources. This phase often surfaces data quality issues, accessibility constraints, and integration challenges that can derail projects before analysis begins. IT leaders recognize this as a critical bottleneck where technical complexity intersects with business requirements.

After this comes data preparation, cleaning, and pipeline construction, which represent the most time-intensive phase, typically consuming 60-80% of total project effort. Data scientists must identify inconsistencies, handle missing values, standardize formats, and create reliable data pipelines. This work requires significant technical expertise while offering limited intellectual satisfaction, a combination that contributes to talent retention challenges.

Data scientists perform exploratory data analysis and hypothesis formation, which involves statistical analysis to understand data characteristics, identify patterns, and formulate testable hypotheses. This phase requires statistical and computational expertise combined with business intuition to generate meaningful analytical directions.

Model Building

In the model development phase, the data scientist takes the prepared data and builds predictive models. This often starts with feature engineering and selection—creating and choosing variables that effectively represent the underlying business problem. These features can be as simple as a ratio or more complex, such as a loyalty score. This process combines statistical knowledge, domain expertise, and creativity, often representing the difference between successful and unsuccessful models.

Next, the data scientist selects the algorithms and determines the model architecture, which requires understanding the strengths and limitations of various analytical approaches. Data scientists must balance model complexity, interpretability requirements, computational constraints, and business objectives. The expanding universe of available algorithms and techniques makes this selection increasingly challenging.

After this, the data scientist trains, validates, and tests the model. This phase involves sophisticated statistical techniques to ensure models generalize beyond training data while avoiding overfitting, a common source of production model failures. The data scientist might fine-tune the model through systematic parameter adjustment. This computationally intensive process requires both statistical knowledge and practical experience to achieve optimal results efficiently.

The model is then registered and version-controlled to ensure reproducibility and enable collaborative development. This often-overlooked aspect becomes critical in enterprise environments where regulatory compliance and audit requirements demand comprehensive documentation.

Operationalization

The deployment phase bridges analytical development with business applications. Model deployment to production environments involves technical integration with existing systems while maintaining performance and reliability standards. This phase often reveals gaps between development and production environments that can compromise model effectiveness. This phase often involves integration with business applications and systems. This integration requires understanding both analytical outputs and business workflows, a combination of skills that creates implementation bottlenecks.

The model must also be validated in the live environment and monitored to maintain model effectiveness over time. Production models face data drift, concept drift, and performance degradation that require continuous monitoring and intervention. Once in production, the model also needs to be maintained to ensure long-term value through regular refresh cycles, performance optimization, and adaptation to changing business requirements.

Governance

Models also need to be governed. This type of governance is different from data governance, and TDWI research indicates that organizations are only getting started with it. Yet, model transparency, explainability, and trust requirements have intensified with regulatory scrutiny and business demands for interpretable AI systems (e.g., the EU AI Act, DORA, etc.). This requires documentation, audit trails, and control requirements that may span the entire model life cycle. It will also include risk management and bias detection to identify potential negative consequences and ensure fair model outcomes across different populations.

This life cycle complexity creates multiple opportunities for AI-infused tools to deliver value while highlighting why traditional approaches struggle to scale effectively across organizations.

> *AI-infused tools augment human capabilities, enabling faster execution while maintaining quality and control.*

AI ENHANCEMENT ACROSS THE LIFE CYCLE

AI-infused tools deliver value by intelligently automating routine tasks, providing expert-level recommendations, and reducing the technical barriers that limit broader organizational participation in data science. Rather than replacing human judgment, these tools augment human capabilities at each life cycle phase, enabling faster execution while maintaining quality and control. TDWI research indicates that these AI-infused automated tools are becoming more important to organizations. For instance, in a recent TDWI survey, the vast majority (78%) of respondents agreed with the statement, "It is important to make use of 'intelligent' tools on the market that incorporate AI/ML for data management for profiling, transforming, and processing data."[1]

These tools help across the four areas of the data science life cycle described earlier: data preparation, model building, model deployment, and governance. Additionally, these tools can help with interactive communication. These benefits are described in the next section.

Enhanced Data Preparation

Automated data profiling transforms the time-intensive process of understanding new data sets. AI tools analyze data structures, statistical distributions, relationships between variables, and quality issues, generating comprehensive reports that might otherwise require hours of manual exploration. These systems identify missing values, outliers, inconsistent formatting, and potential data quality problems while providing statistical summaries and visualizations that accelerate initial data assessment.

---

[1] Unpublished 2025 TDWI Data and Analytics survey.

Smart data exploration introduces natural language interfaces that enable business analysts and domain experts to participate directly in data discovery. Rather than requiring SQL or programming expertise, users can make requests like "show me sales trends by region" or "identify unusual customer behavior patterns" and receive relevant analyses. This capability democratizes data access while maintaining analytical rigor through AI-powered query interpretation and execution.

Intelligent transformation recommendations leverage machine learning to suggest appropriate data cleaning and preparation steps based on identified patterns, data types, and quality issues. These systems learn from successful transformation patterns across similar data sets, recommending standardization approaches, missing value handling strategies, and data type conversions that optimize downstream analysis effectiveness.

Anomaly detection capabilities automatically identify outliers, inconsistencies, and unusual patterns that might indicate data quality problems or interesting business phenomena. This automation catches issues that manual review might miss while flagging potential insights that warrant further investigation.

> *AutoML enables less experienced analysts to build sophisticated models while freeing expert data scientists to focus on complex problem-solving and business strategy.*

Accelerated Model Building

AutoML integration represents potentially the most transformative enhancement, automating algorithm selection, feature engineering, and hyperparameter optimization processes that typically require significant expertise and experimentation. These systems evaluate multiple modeling approaches, automatically create and test feature combinations, and optimize model parameters to achieve optimal performance for specific business objectives. This automation enables less experienced analysts to build sophisticated models while freeing expert data scientists to focus on complex problem-solving and business strategy.

Intelligent code assistance provides real-time support for R and Python development through context-aware suggestions, code completion, and error detection. These tools understand analytical workflows, suggesting relevant functions, identifying potential bugs, and providing explanations for complex code segments. This capability accelerates development while serving as a learning platform for analysts expanding their technical skills.

Model architecture recommendations analyze problem characteristics, data properties, and performance requirements to suggest optimal modeling approaches. Rather than manual experimentation with different algorithms, these systems provide intelligent starting points based on similar successful projects and theoretical best practices.

Rapid prototyping tools lower barriers for analysts new to advanced modeling techniques by providing template-based approaches with intelligent customization. These platforms enable quick model development and testing cycles, allowing organizations to evaluate analytical approaches before committing significant resources to full development.

Streamlined Deployment and Operations

Automated monitoring systems continuously evaluate model performance, data quality, and business impact in production environments. These tools detect data drift (changes in input data characteristics over time), concept drift (changes in underlying relationships between variables), and performance degradation that can compromise model effectiveness. Automated alerts ensure prompt intervention when models require attention or retraining.

Intelligent scheduling optimizes computational resources and job execution timing based on system load, data availability, and business requirements. This automation ensures efficient resource utilization while maintaining service-level agreements for critical analytical processes.

Version management capabilities provide AI-assisted model versioning, comparison, and rollback functionality. These systems track model lineage, performance metrics, and deployment history, enabling rapid response to production issues while maintaining comprehensive audit trails.

Interactive Communication

Visualization and natural language generation capabilities automatically create dashboards, reports, and stakeholder communications that translate technical analysis into business-relevant insights. These tools generate narrative explanations of analytical findings, create appropriate visualizations for different audiences, and maintain consistent reporting standards across projects. This automation addresses the communication gap between technical teams and business stakeholders while ensuring analytical insights drive business action.

Enhanced Governance and Compliance

Automated documentation generates comprehensive model documentation, metadata management, and technical specifications required for regulatory compliance and knowledge transfer. These systems capture model assumptions, data sources, transformation logic, and performance characteristics in standardized formats that support audit requirements and collaborative development.

Bias detection tools automatically evaluate models for potential discriminatory outcomes across different population segments. These systems test for statistical bias, fairness metrics, and disparate impact while providing recommendations for bias mitigation strategies. This capability addresses growing regulatory requirements while supporting ethical AI deployment.

Lineage tracking provides automated data and model lineage documentation that traces analytical processes from source data through final outputs. This capability supports reproducibility requirements, impact analysis, and compliance documentation while enabling efficient troubleshooting and model debugging.

Alert systems provide intelligent monitoring that triggers notifications when models require attention, compliance issues arise, or performance thresholds are exceeded. These systems prioritize alerts based on business impact and provide relevant context for rapid problem resolution.

The cumulative effect of these enhancements can help transform data science from a specialized, resource-intensive discipline into a more accessible, scalable organizational capability while maintaining professional standards and governance requirements.

Critical Plays for Getting Started

Successful AI-infused data science implementation requires systematic planning, strategic tool selection, and organizational change management. The following five critical plays provide a structured approach for IT leaders to navigate this transformation while minimizing risk and maximizing business value. These plays are designed to be executed in sequence, with each building upon the foundation established by the previous play, though organizations may adapt the timeline and emphasis based on their specific maturity level and business requirements.

PLAY 1: ASSESSING TEAM MATURITY AND DEVELOPING AN INTEGRATION STRATEGY

Implementing AI-infused tools within the data science life cycle requires an evaluation of the current team's capabilities and the development of a strategic integration plan.

Begin by conducting a comprehensive skills inventory to assess proficiency levels in languages such as Python and R, as well as expertise in statistical modeling. Utilize standardized assessment tools to map individual competencies against project requirements and future organizational needs. This process will help identify strengths, gaps, and areas requiring development.

Understanding the team's familiarity with both commercial and open source tools is also important. Survey team members to gauge their experiences and preferences regarding existing data science platforms. Evaluate current tool utilization rates and identify any resistance to adopting new technologies.

Mapping existing workflows is important for pinpointing bottlenecks and inefficiencies. Document the current processes from problem definition through deployment, highlighting manual handoffs and time-intensive activities. This analysis will help inform the strategic planning components of AI integration.

Define clear objectives for incorporating AI, setting measurable goals such as reducing routine tasks by 30-50%. Align these objectives with broader digital transformation initiatives and establish success metrics to monitor progress. Metrics are critical for success because they help illustrate progress and successes. Some organizations will implement pilot projects to demonstrate quick wins, ensuring that each phase has clear evaluation criteria and risk mitigation strategies.

Develop a compelling business case by projecting ROI through cost savings, efficiency gains, and enhanced decision-making capabilities. Consider implementation costs, including licensing, training, and change management, to provide a comprehensive financial overview.

PLAY 2: PRIORITIZING TOOLS WITH EMBEDDED INTELLIGENCE AND HUMAN OVERSIGHT

Successful AI integration in data science requires selecting tools that enhance human capabilities while maintaining appropriate oversight. The key lies in choosing platforms that provide intelligent assistance without compromising the critical thinking and domain expertise that analysts bring to complex problems.

Essential AI-enhanced features should include sophisticated code assistance with context-aware completion for Python and R environments, real-time error detection, and integrated documentation lookup. Natural language interfaces can democratize data exploration by enabling business stakeholders to translate questions directly into analytical code, while automated insight generation through pattern recognition accelerates discovery and ensures significant findings aren't overlooked.

Template libraries with AI-guided customization provide foundations for consistent, high-quality work. Industry-specific templates addressing common business problems reduce project start-up time, while intelligent selection based on data characteristics ensures appropriate methodological approaches.

However, preserving human decision-making authority remains paramount. Approval workflows for model deployment and strategic recommendations ensure human expertise stays central to the analytical

process. AI suggestions should require explicit human acceptance rather than automatic implementation, with clear escalation paths for complex scenarios exceeding AI capabilities.

Transparency becomes essential for maintaining trust and enabling effective human-AI collaboration. Clear attribution of AI-generated versus human-created content, combined with explanations for recommendations, helps users understand and evaluate assistance received. Comprehensive audit trails support both compliance requirements and continuous improvement.

Finally, validation checkpoints throughout automated processes provide essential quality controls. Automated testing for AI-generated code, business logic validation against known constraints, and robust exception handling ensure reliability while highlighting areas where human intervention remains necessary for optimal outcomes.

PLAY 3: ENCOURAGING A CULTURE OF SAFE EXPLORATION AND INNOVATION

Enabling innovation with AI tools requires creating environments where teams can experiment freely while maintaining appropriate security and governance controls. The challenge lies in balancing exploration opportunities with organizational risk management and data protection requirements.

Establishing dedicated sandbox environments provides the foundation for safe experimentation. These isolated environments should offer full AI tool access with restricted data and system connectivity, enabling meaningful experimentation without production impact. Automated environment provisioning allows rapid setup for new experiments, while shared sandbox spaces facilitate collaborative learning and knowledge transfer across teams.

Data protection strategies become critical for maintaining security during experimentation. Data masking tools that preserve analytical characteristics while protecting sensitive information enable realistic testing scenarios.

Synthetic data sets that maintain statistical properties of production data provide additional experimentation opportunities, while clear protocols for data movement between environments ensure appropriate governance controls remain in place.

Reducing friction while maintaining control requires implementing user-friendly interfaces that democratize data access. Natural language querying tools that translate business questions into database queries enable broader organizational participation in analytical activities. Template notebooks with AI-assisted customization provide proven starting points for new projects, while intelligent template matching suggests relevant approaches based on project characteristics.

Creating shared knowledge repositories captures organizational learning and accelerates capability development. Searchable repositories of successful workflows, code samples, and lessons learned enable teams to build on previous successes rather than starting from scratch. Communities of practice provide forums for sharing experiences and problem-solving approaches, while mentorship programs pair experienced users with newcomers.

Systematic documentation through experiment tracking systems ensures that insights are captured and shared. Regular review sessions to discuss results and lessons learned, combined with feedback loops that inform future tool selection, create continuous improvement cycles that enhance organizational AI capabilities over time.

PLAY 4: IMPLEMENTING COMPREHENSIVE DATA AND AI LITERACY TRAINING

Successful AI integration requires comprehensive training programs that develop both technical proficiency and critical thinking skills. Organizations must invest in upskilling existing talent while building new capabilities that enable teams to leverage AI tools effectively and responsibly.

Technical skills development should begin with AI tool proficiency training tailored to existing Python and R users. Hands-on programs that integrate AI assistance into familiar workflows help experienced analysts adapt their

expertise to enhanced capabilities. Proficiency pathways that advance users from basic assistance to advanced automation provide structured learning progressions, while practical exercises using organizational data ensure relevance and immediate applicability.

Traditional analysts require specialized bridge training that connects familiar analytical concepts with AI-enhanced approaches. Learning paths should gradually introduce AI assistance without overwhelming users, supported by mentoring programs that pair experienced AI users with traditional analysts. Low-stakes practice projects build confidence while demonstrating practical value, encouraging broader adoption across analytical teams.

> *Advanced training for power users should focus on prompt engineering and optimization techniques that maximize AI tool effectiveness.*

Advanced training for power users focuses on prompt engineering and optimization techniques that maximize AI tool effectiveness. Specialized instruction on combining multiple AI tools for complex workflows, along with training on AI limitations and failure modes, ensures sophisticated users can navigate edge cases and system boundaries effectively.

Equally important is developing AI literacy and critical thinking capabilities. Users must understand AI capabilities and limitations to set realistic expectations and avoid over-reliance or under-utilization. Training should focus on recognizing when to rely on versus override AI recommendations, providing frameworks for evaluating suggestion quality and appropriateness.

Validation skills become essential for maintaining analytical rigor. Teams need systematic approaches for testing AI-generated code and insights, combining statistical validation with business logic checking. Building awareness of AI bias and fairness considerations ensures responsible use, while establishing organizational standards for bias detection and mitigation protects against unintended consequences that could compromise analytical integrity or organizational reputation.

PLAY 5: ESTABLISHING ROBUST SECURITY AND GOVERNANCE FRAMEWORKS

Implementing AI-enhanced data science capabilities requires comprehensive security and governance frameworks that protect organizational assets while enabling innovation. The challenge lies in creating structures that provide appropriate oversight without stifling the experimentation and agility that AI tools can deliver.

Security considerations must address the unique risks associated with cloud-based AI services and external tool integration. Data classification systems should determine appropriate AI service usage levels, while data residency requirements and vendor compliance verification ensure regulatory alignment. Encryption standards for data transmission and storage, combined with comprehensive access logging and monitoring, provide foundational security controls that protect sensitive information throughout AI-assisted workflows.

Comprehensive guidelines that balance innovation with risk management ensure controlled expansion of AI capabilities.

Secure API management becomes critical for controlling access to AI services while maintaining operational efficiency. API gateways that monitor and control service access, integrated with organizational identity management systems, ensure appropriate authentication and authorization. Rate limiting and usage monitoring prevent unauthorized consumption, while API key rotation and incident response procedures address security vulnerabilities.

Intellectual property protection requires careful consideration of proprietary models and data sharing with AI services. Contractual protections with service providers regarding data usage and model training, combined with internal policies governing proprietary algorithm sharing, help maintain competitive advantages while enabling AI assistance.

Governance infrastructure must establish clear policies for AI tool usage across different data classification levels. Comprehensive guidelines that balance innovation with risk management, supported by approval workflows for new tool adoption, ensure controlled expansion of AI capabilities. Audit

trails that capture AI assistance throughout analytical processes provide transparency and support regulatory compliance requirements.

Model documentation and explainability requirements become essential for maintaining analytical rigor and stakeholder confidence. Documentation standards that capture AI tool usage in development and deployment, combined with explainability tools and validation procedures, ensure AI-assisted models remain interpretable and reliable.

Regular governance reviews and incident response procedures complete the framework, ensuring continuous improvement and effective management of AI-related challenges that may emerge as capabilities evolve.

Conclusion

The convergence of artificial intelligence with open source data science represents a transformational opportunity for organizations to scale analytical capabilities while managing complexity and cost. AI-infused tools resolve the fundamental tension between open source's innovation advantages and enterprise operational requirements, enabling broader organizational participation in data science without compromising professional standards or governance controls.

For IT leaders, this evolution addresses persistent challenges that have limited AI adoption velocity: skills gaps, resource constraints, governance complexity, and integration difficulties. By intelligently automating routine tasks while preserving human decision-making authority, AI-infused platforms democratize advanced analytics capabilities across the organization while maintaining the flexibility and cost advantages that make open source tools attractive.

The business case extends beyond efficiency gains to encompass strategic transformation. Organizations implementing AI-infused data science can expect faster time-to-insight, improved model quality through automated best practices, reduced operational risk through enhanced governance capabilities, and expanded analytical capacity without proportional increases in specialized talent requirements. These benefits compound over time as teams become more proficient with AI assistance and organizational analytical maturity increases.

However, success requires systematic implementation guided by a clear strategic vision. The five critical plays outlined (team assessment and strategy development, intelligent tool selection with human oversight, culture change management, comprehensive training programs, and robust governance frameworks) provide a structured approach for navigating this transformation while minimizing implementation risks.

> *The question is not whether AI will transform data science; that transformation is already underway.*

The window for competitive advantage through AI-infused data science is narrowing as these capabilities become more accessible and adoption accelerates across industries. Organizations that move decisively to implement these technologies while their competitors struggle with traditional data science constraints will establish analytical capabilities that become increasingly difficult to replicate.

The path forward demands balancing innovation with prudent risk management, embracing automation while preserving human expertise, and scaling capabilities while maintaining quality standards. IT leaders who successfully navigate these tensions will position their organizations to leverage data science as a sustainable competitive advantage rather than a specialized, resource-intensive capability.

The question is not whether AI will transform data science; that transformation is already underway. The critical decision facing IT leaders is whether to lead this transformation within their organizations or react to competitive pressures after others have established analytical advantages. The tools, frameworks, and strategies exist today to begin this journey successfully.

Better Together: Posit and Databricks

(Content supplied by Posit)

Posit PBC, a leader in open source and professional software for data science teams using Python and R, is seamlessly integrated with the Databricks Data Intelligence Platform and was recently awarded the Databricks Developer Tools Partner of the Year. Customers can leverage this essential partnership to build, deploy, and share insights faster—using the tools they know and the data platform they trust.

Over the past year, Posit has demonstrated an unparalleled commitment to helping enterprise organizations build data intelligence through its collaboration with Databricks, "the Data and AI company." Posit provides tools that simplify and streamline the data science life cycle—from development and collaboration to deployment and management.

Many organizations face challenges in bridging the gap between powerful open source tooling favored by data scientists and the enterprise requirements for governance, security, and scalability, especially when working with large data and AI platforms like Databricks. Posit's solutions, including Posit Workbench, Posit Connect, and Posit Package Manager, address these challenges by offering a professional, integrated environment for R and Python practitioners to work directly with data resources managed by Databricks. Together, Posit and Databricks are working to ensure that data scientists can easily connect to Databricks, browse data, and build models and applications from within their favorite editors like RStudio and Positron.

Posit and Databricks solve the challenge facing enterprise data science teams today: the ability to develop quickly using their preferred tools amidst workflow fragmentation and management complexity, all while maintaining the governance that their organizations need. Our joint solution provides faster and more secure data science development.

Learn more and get started.

About Our Sponsors

Databricks is the Data and AI company. More than 15,000 organizations worldwide—including Block, Comcast, Condé Nast, Rivian, Shell, and over 60% of the *Fortune* 500—rely on the Databricks Data Intelligence Platform to take control of their data and put it to work with AI. Databricks is headquartered in San Francisco with offices around the globe and was founded by the original creators of Lakehouse, Apache Spark, Delta Lake, MLflow, and Unity Catalog.

Posit is a Public Benefit Corporation and a certified B Corporation, dedicated to creating open-source and commercial software for data science, scientific research, and technical communication. As the creators of RStudio, Shiny, the Tidyverse, and other open source software for data science, Posit develops and supports tools that allow individuals and teams to easily create, share, and consume knowledge using Python and R.

About the Authors

Donald Farmer is an internationally respected speaker and writer with over 30 years of experience in data management and analytics. He advises investors, software vendors, and enterprises on data and analytics strategy. Donald led teams at Microsoft and Qlik, building some of the most important and innovative analytics platforms on the market. His consultancy, TreeHive Strategy, has developed a process known as the Gameplan, which enables enterprises to map out an effective, step-by-step strategy for developing a culture of analytics and data literacy. In addition to data strategy, Donald frequently consults on governance and compliance issues, particularly for companies implementing self-service architectures. His workshops on innovation strategy have helped to drive significant changes in numerous businesses.

Fern Halper, Ph.D., is vice president and senior director of TDWI Research for advanced analytics. She is well known in the analytics community, having been published hundreds of times on data mining and information technology over the past 20 years. Halper is also coauthor of several Dummies books on cloud computing and big data. She focuses on advanced analytics, including predictive analytics, machine learning, AI, cognitive computing, and big data analytics approaches. She has been a partner at industry analyst firm Hurwitz & Associates and a lead data analyst for Bell Labs. She has taught at both Colgate University and Bentley University. Her Ph.D. is from Texas A&M University.

You can reach her by email (fhalper@tdwi.org) and on LinkedIn (linkedin.com/in/fbhalper).

About TDWI Research

TDWI Research provides industry-leading research and advice for data and analytics professionals worldwide. TDWI Research focuses on modern data management, analytics, and data science approaches and teams up with industry thought leaders and practitioners to deliver both broad and deep understanding of business and technical challenges surrounding the deployment and use of data and analytics. TDWI Research offers in-depth research reports, commentary, assessment, inquiry services, and topical conferences as well as strategic planning services to user and vendor organizations.

About TDWI Playbooks

TDWI Playbooks provide data professionals with a summary of important key factors about contemporary data-related topics. Playbooks present the issues and challenges facing enterprises about each topic and offer a concise list of proven best practices to succeed in a particular area of analytics, business intelligence, or data management. Playbooks are written by TDWI research analysts and faculty who synthesize their research and experience into easy-to- understand explanations and practical recommendations that enable data professionals to apply the best, most productive approaches and techniques to their projects or initiatives.

tdwi

**Transforming Data
With Intelligence™**

A Division of 1105 Media
6300 Canoga Avenue, Suite 1150
Woodland Hills, CA 91367

E info@tdwi.org

tdwi.org