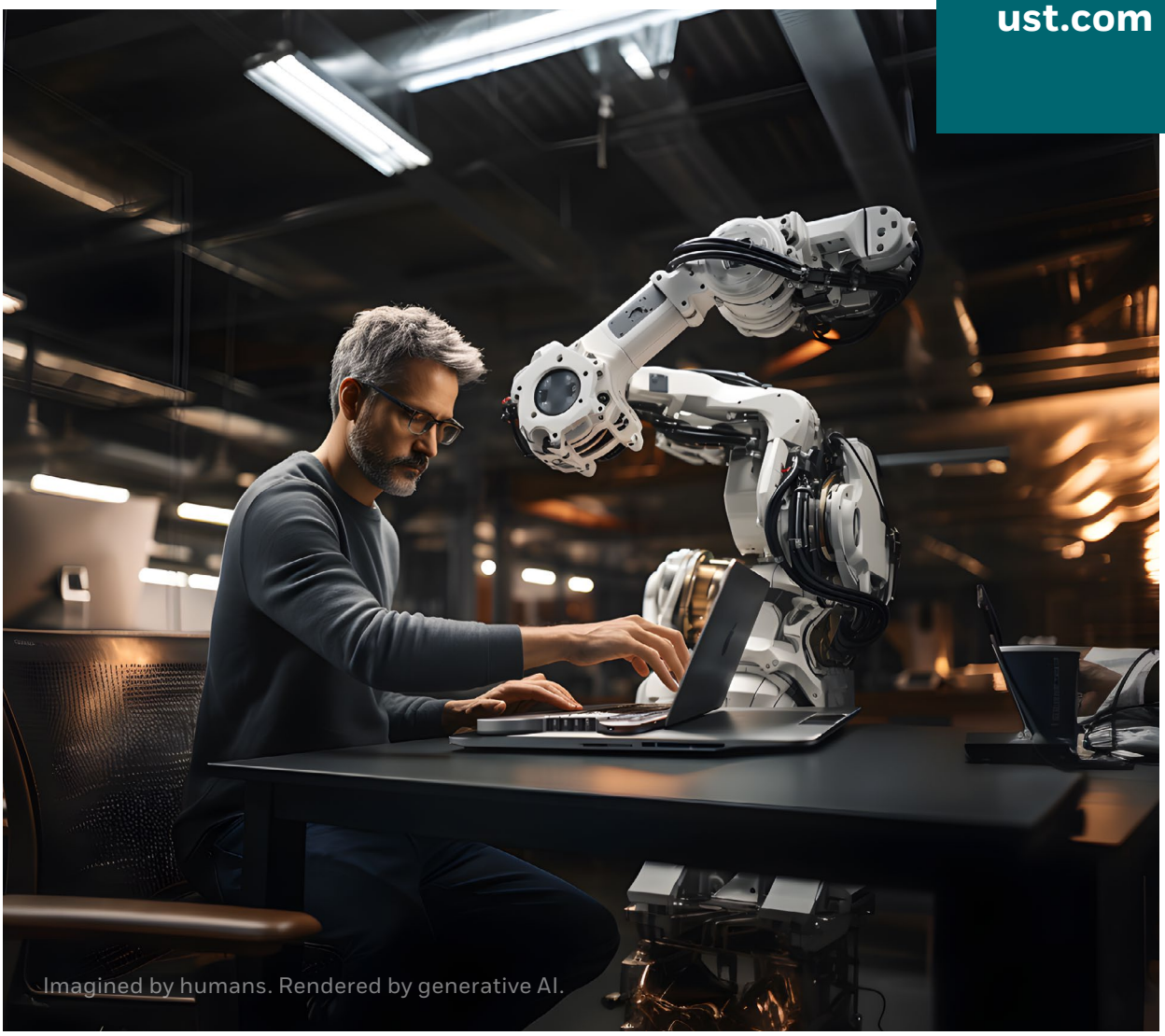
Trust and boundaries in Agentic AI

WHITEPAPER

When to trust Agentic AI — and when not to

ust.com



Imagined by humans. Rendered by generative AI.

UST

Table of contents

ABSTRACT

Agentic AI promises intelligence with initiative, but unchecked autonomy can lead to ethical, legal, and operational pitfalls. This whitepaper provides a strategic framework for identifying the right moments to delegate decision-making to AI—and when to ensure human control remains essential.

RESOURCES

1. https://www.ust.com/en/ust-explainers/agentic-ai-the-next-frontier-in-artificial-intelligence
2. https://www.ust.com/en/smartops/platform
3. https://artificialintelligenceact.eu/article/14/
4. https://standards.ieee.org/initiatives/autonomous-intelligence-systems/standards/
5. https://www.nist.gov/itl/ai-risk-management-framework
6. https://www.iso.org/standard/74438.html

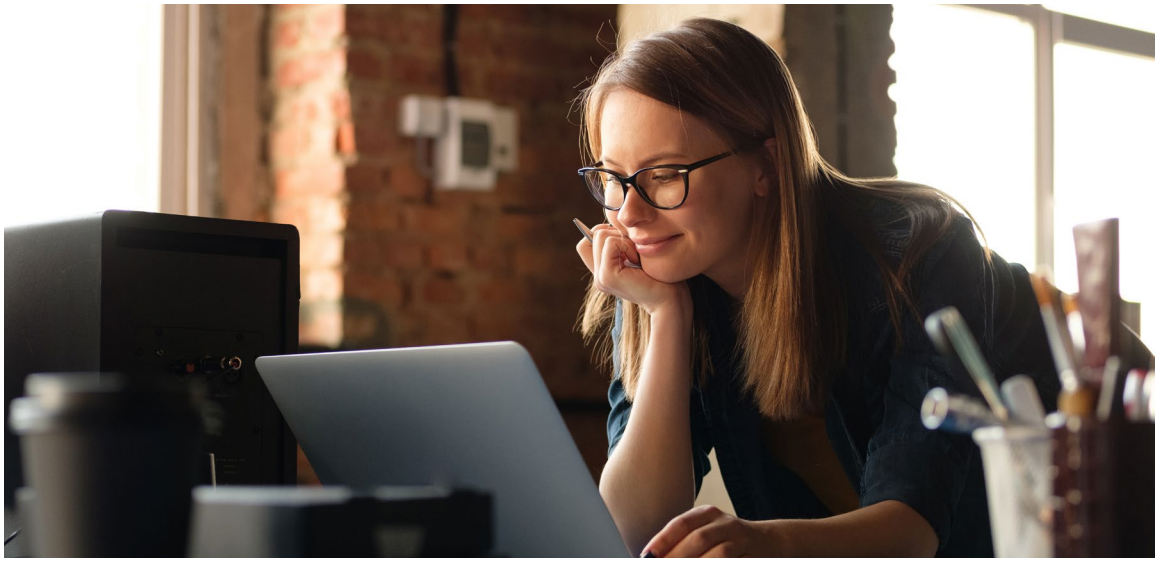
Introduction

Agentic AI refers to artificial intelligence systems designed to operate with genuine autonomy: sensing, reasoning, planning, and acting in real-world, often multi-agent environments. Unlike traditional AI that follows predefined rules or scripts, Agentic AI systems dynamically deconstruct complex goals into subtasks, adapt strategies in response to changing contexts, and learn from experience without continuous human oversight.

Examples include autonomous vehicles navigating unpredictable traffic and algorithmic trading agents making real-time investment decisions. What sets Agentic AI apart is its initiative-driven behavior: systems that don't just react, but proactively decide, learn, and collaborate. This level of autonomy requires cognitive capabilities such as goal planning, memory, inter-agent coordination, and real-time adaptation especially crucial in uncertain and high-stakes domains.

The real problem despite growing interest and rapid progress in Agentic AI, a fundamental challenge remains: **when should we trust autonomous agents—and when should we not?** Much of the current discourse focuses on technical performance, speed, accuracy, and efficiency, often at the expense of safety, human context, ethical accountability, and socioeconomic impact.

The lack of clear decision boundaries where Agentic AI can safely operate without human input raises significant concerns. Critical questions remain unanswered: In which domains do the risks of autonomy outweigh the benefits? When should systems yield control to humans? What governance mechanisms must be in place to ensure responsible autonomy at scale?



Defining agentic AI

Agentic AI systems differ from traditional AI in that they act with autonomy, not merely on fixed inputs, but in anticipation of complex and evolving scenarios. These agents can adapt to novel conditions and optimizing outcomes using real-time signals operating beyond the rigid confines of deterministic rule sets.

By contrast, non-agentic systems require explicit instructions for each task and lack the capacity for autonomous reasoning or situational awareness. While well-suited for repetitive, structured processes, such systems falter in fluid environments where adaptation is critical.

GARTNER PREDICTS OVER 40% OF AGENTIC AI PROJECTS WILL BE CANCELLED BY END OF 2027

Read the article →

Agentic AI systems

Decoding intent

Understanding commands in natural language, reducing dependency on static menus

Enhancing accessibility

Offering intuitive, adaptive interface that require minimal users training

Autonomous decision-making

Generating contextual responses ad evolving strategies based on user goals and environmental cues

Example

An autonomous AI agent that can intelligently adapt its dialogue and problem-solving approach-evolving its strategies in real-time to meet complex user needs.

Non-Agentic systems

Depend on manual control

Require constant user navigation and input through fixed GUIs

Lack adaptability

Offer limited flexibility, often becoming cumbersome in multi-step or complex operations

Cause user frustration

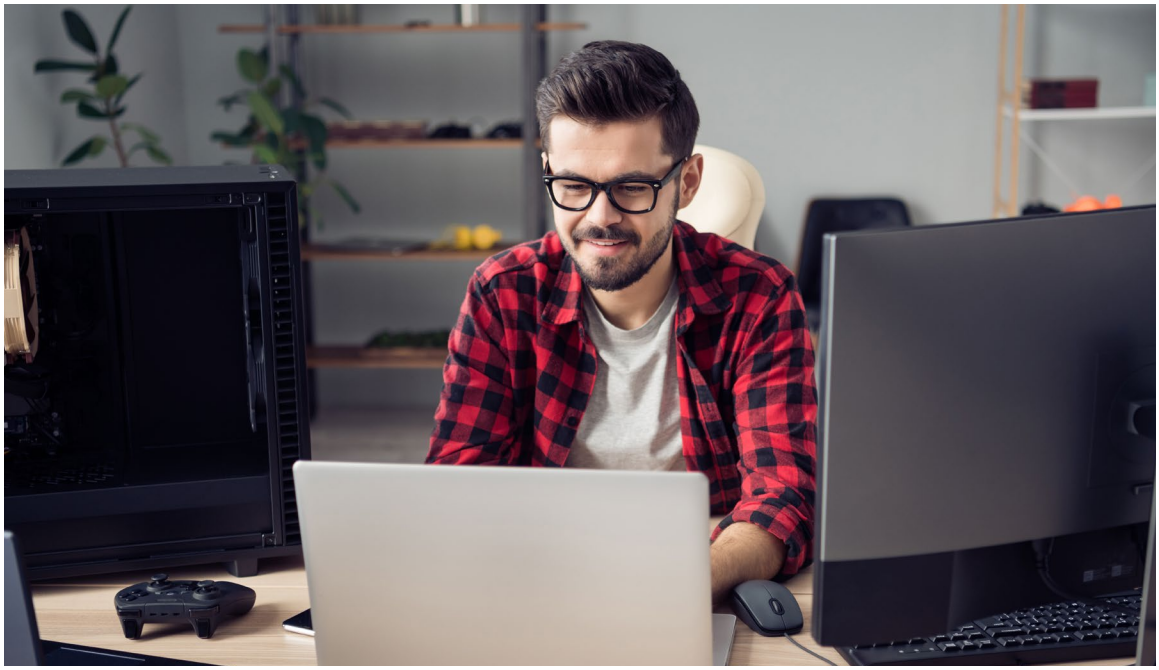
Fail to anticipate user needs, leading to inefficiency and dissatisfaction

Example

A conventional chatbot that relies solely on scripted responses, unable to move beyond programmed logic or contextual variance.

This distinction underscores the strategic value of Agentic AI in dynamic, data-intensive settings but also reinforces the urgency of defining clear boundaries for its trusted use.

When to use agentic AI



A structured framework for deploying Agentic AI can help organizations determine when autonomous decision-making is appropriate. Key factors include task complexity, adaptability requirements, efficiency gains, ethical impact, and organizational readiness. With this lens, organizations can identify contexts where Agentic AI enhances performance, mitigates risk, and operates responsibly especially in dynamic or high-stakes environments.

HIGH-COMPLEXITY TASKS

Agentic AI proves most effective in high-complexity tasks where conventional automation lacks the flexibility or foresight to function reliably. These are environments characterized by rapid variable shifts, uncertainty, and a constant need for multi-step decision-making. In such scenarios, the ability of Agentic AI to decompose goals, continuously interpret evolving input, and autonomously orchestrate decisions enables operational continuity. Traditional deterministic systems, which depend on fixed logic, fail to deliver under these conditions due to their inability to generalize or respond to unforeseen dynamics.

TASKS REQUIRING CONSTANT ADAPTABILITY

Certain domains demand real-time responsiveness and adaptive reasoning to navigate variability. When operational environments are exposed to continual change be it external disruptions, fluctuating constraints, or interdependencies Agentic AI's dynamic behavior becomes a strategic asset. Its capacity to absorb feedback, recalibrate priorities, and optimize execution paths

enables improved agility and reduced latency. In these settings, static rule sets quickly become obsolete, while adaptive agents sustain performance through continuous alignment with shifting goals.

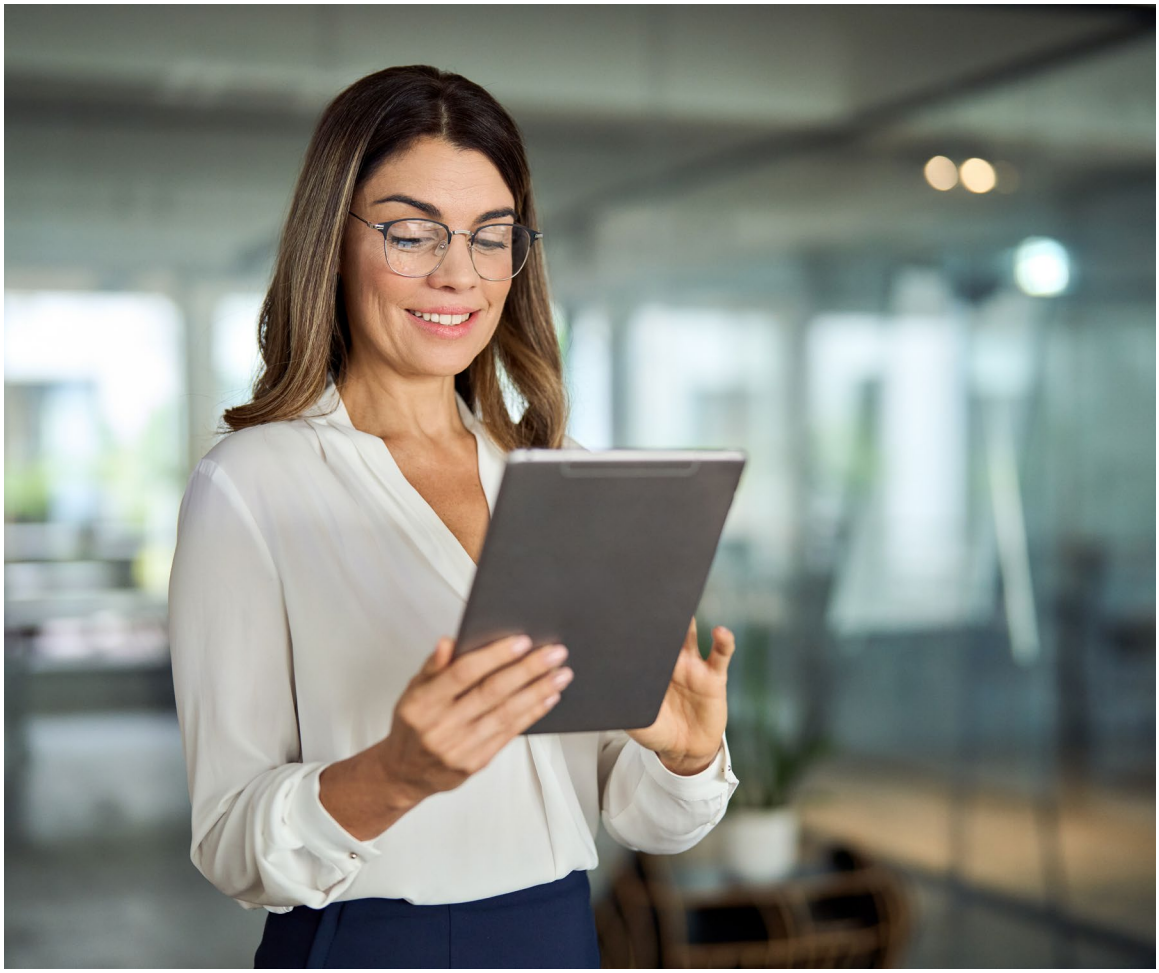
EFFICIENCY GAINS WITH MINIMAL RISK

Agentic AI also holds significant potential in scenarios where performance can be enhanced without introducing disproportionate risks. These are typically well-instrumented environments where failure modes are understood, control boundaries are defined, and autonomous intervention is supported by robust oversight mechanisms. Here, agentic systems can forecast operational needs, fine-tune resource deployment, and manage multi-threaded activities with limited human input. The result is measurable improvement in efficiency, reliability, and cost-effectiveness especially when autonomy is augmented with monitoring and fallback protocols.

ETHICAL CONSIDERATIONS

As autonomy increases, so does the ethical responsibility of those deploying it. Agentic AI must not only perform tasks effectively but also do so in a manner that is fair, transparent, and accountable. Without appropriate constraints, these systems can unintentionally reinforce bias, compromise privacy, or escalate systemic risk. Ethical implementation requires explicit design for human-in-the-loop supervision, algorithmic auditability, and safeguards that reflect both societal norms and regulatory mandates. High-stakes applications particularly those impacting public welfare or individual rights demand an even higher threshold of justification, oversight, and governance.

When not to use agentic AI



While Agentic AI offers transformative potential, its application is not universally appropriate. There are domains where its use introduces unnecessary complexity, increases risk, or conflicts with foundational principles of safety, accountability, and ethics. In such contexts, simpler automation or human decision-making remains more effective and responsible.

LOW COMPLEXITY OR REPETITIVE TASKS

In operational settings defined by predictability and routine, Agentic AI often adds unwarranted overhead. Tasks that follow fixed sequences, require no contextual interpretation, or involve minimal variance are better suited to deterministic automation or traditional scripting. Introducing agentic systems in these cases not only increases deployment and maintenance costs but also risks reducing reliability due to unnecessary abstraction. In such environments, simplicity ensures speed, traceability, and cost-efficiency outcomes that autonomy cannot meaningfully improve.

> Autonomy is powerful—but not always appropriate. In high-stakes domains, AI must pause for permission before taking the next step.

HIGH RISK OR LIFE SENSITIVE ENVIRONMENTS

Domains involving life-critical decisions or irreversible consequences demand extreme caution. Agentic systems, by their nature, may evolve strategies, introduce novel behaviour, or make decisions that are difficult to interpret in real time. In healthcare, aviation, nuclear operations, or military strategy, even marginal errors can have catastrophic implications. In such environments, the inability to predict or fully audit autonomous reasoning undermines the essential trust required for deployment. Human oversight and clearly bounded automation remain indispensable where stakes are high and tolerance for uncertainty is minimal.

LACK OF ACCOUNTABILITY OR TRANSPARENCY

Agentic AI should not be deployed in systems where decision logic cannot be interrogated, or responsibility cannot be clearly assigned. The opacity of certain models, particularly those based on deep learning, presents significant challenges for auditing and governance. When actions taken by an agent affect financial transactions, legal determinations, or regulatory compliance, the absence of explainability exposes institutions to reputational, legal, and systemic risks. Transparent decision-making and traceable reasoning are prerequisites in contexts that demand accountability, and any system lacking these qualities is fundamentally unfit for autonomous operation.

ETHICAL OR SOCIETAL CONCERNS

Autonomous systems can inadvertently amplify societal harms if deployed without strong ethical safeguards. The misuse of Agentic AI in areas such as surveillance, information manipulation, or mass behavioural influence has already raised serious concerns. When systems act independently without human-centered constraints, they may encode bias, infringe on privacy, or destabilize social trust. As the scale and reach of these systems grow, so too does their potential for disproportionate harm. For Agentic AI to be socially viable, its design and deployment must align with principles of fairness, transparency, legal compliance, and respect for human dignity.

Risk management and mitigation strategies

IMPORTANCE OF RESPONSIBLE DEPLOYMENT

As Agentic AI systems extend into high-impact domains such as healthcare, finance, and transportation, responsible deployment becomes essential. Unlike conventional AI, these systems operate with autonomy making decisions and initiating actions in dynamic contexts. While powerful, this autonomy introduces risks around safety, bias, accountability, and system drift. Without robust safeguards, these systems can cause unintended harm or ethical breaches.

To address this, regulatory and industry frameworks now emphasize human oversight, auditability, and explainability. The EU AI Act, Article 14 exemplifies this direction, enforcing

risk-based governance for high-stakes applications. Responsible deployment ensures Agentic AI enhances human well-being, complies with societal norms, and earns long-term trust.

MONITORING AND CONTROL MECHANISMS

Maintaining control over Agentic AI requires integrated monitoring strategies. Human-in-the-loop (HITL) systems allow AI to propose actions while reserving execution for human approval essential in regulated or sensitive domains. In parallel, automated monitoring through real-time validation, anomaly detection, and system logging enables scalable oversight.

These mechanisms ensure that autonomy remains bounded, verifiable, and correctable in critical moments

TRANSPARENCY AND INTERPRETABILITY

Agentic systems must offer transparent reasoning to support trust and accountability. Techniques such as thought tracing and prompt logging help interpret complex decisions, while system documentation clarifies purpose, data sources, and constraints. User-facing disclosures and audit trails further reinforce explainability across the AI lifecycle.

Together, these tools ensure that agentic behaviour remains intelligible, defensible, and compliant.

> “
>
> Trust in AI isn’t built on code alone—it requires ethics, oversight, and transparent guardrails.

ETHICAL OVERSIGHT AND GOVERNANCE

Ethical deployment requires more than technical controls—it demands structured governance. Regulatory frameworks like the EU AI Act, alongside standards such as IEEE 7000 and NIST’s AI RMF, provide a blueprint for aligning autonomy with societal values.

Organizational governance includes internal review boards, third-party audits, and adherence to ethical design principles. Pre-deployment assessments, compliance protocols, and continuous oversight ensure Agentic AI operates within legal, ethical, and human-aligned boundaries.

Recommendations for organizations

Organizations exploring Agentic AI should begin by evaluating its alignment with their operational goals, complexity of tasks, and tolerance for autonomous decision-making. Agentic AI is best suited for dynamic, high-risk environments but not all processes benefit equally from autonomy. Prioritizing use cases where Agentic AI outperforms traditional automation is key.

Ethical, legal, and regulatory compliance must be embedded from the outset. Align implementations with evolving frameworks such as the EU AI Act, IEEE 7000, and ISO/IEC 23053 to ensure transparency, accountability, and auditability. Establishing cross-functional governance teams including experts in data science, law, cybersecurity, and ethics strengthens oversight, mitigates risk, and supports responsible lifecycle management.

Conclusion

Agentic AI holds transformative potential across industries by enabling autonomous, context-aware decision-making in complex domains. Real-world outcomes such as improved claims processing, agile supply chain responses, and legacy modernization demonstrate its measurable impact.

However, the risks are real. Challenges such as inconsistent data quality, lack of domain-specific knowledge, and limited regulatory guidance can undermine even the most promising implementations. Success requires more than technical innovation it demands human oversight, ethical design, and proactive governance. When approached with responsibility and foresight, Agentic AI can significantly enhance human capability while ensuring transparency, fairness, and trust.

Author

Dr. Renjith Paulose

Principal AI Architect
UST SmartOps
Kochi, India
renjith.paulose@ust.com

Together, we build for boundless impact

Since 1999, UST has worked side by side with the world's best companies to make a powerful impact through transformation. Powered by technology, inspired by people, and led by our purpose, we partner with our clients from design to operation. Our digital solutions, proprietary platforms, engineering, R&D, products, and innovation ecosystem turn core challenges into impactful, disruptive solutions. With deep industry knowledge and a future-ready mindset, we infuse expertise, innovation, and agility into our clients' organizations—delivering measurable value and positive lasting change for them, their customers, and communities around the world. Together, with 30,000+ employees in 30+ countries, we build for boundless impact—touching billions of lives in the process.

ust.com

Version 0104-20251027

U ▪
S T