ADGM Academy Research Centre

أكاديمية ربدان
Rabdan Academy

standard chartered

Navigating the AI Frontier: Ensuring Human-Centric Innovation in a Tech-Centric World

AUTHORED BY

- Dr. John George Hatzadony, Assistant Professor – Programme Chair Homeland Security, Rabdan Academy

WITH CONTRIBUTIONS BY:

- Emma Johnson, Executive Director, AI Safety and AI Talent, Standard Chartered Bank
- Azra Razak, Director, AI Safety, Group Chief Data Office, Standard Chartered Bank
- Farooq Husain, Director, AI Safety, Group Chief Data Office, Standard Chartered Bank

Abstract

As artificial intelligence transforms industries at breakneck speed, we're witnessing a fundamental debate about how to implement these powerful tools. This paper explores the growing tension between human-centric and tech-centric approaches to AI, making the case that artificial intelligence works best when it amplifies human capabilities rather than substituting for them. We draw on Garry Kasparov's groundbreaking insights about human-machine collaboration in chess, the U.S. military's innovative digital centaur framework, and cutting-edge research from leading institutions to outline practical strategies for keeping AI tools under human direction while unleashing their full potential to improve decision-making across industries. Through detailed analysis of spectacular AI failures—from legal practitioners facing hefty sanctions to discriminatory lending algorithms—we examine what happens when organisations rely too heavily on automated systems. The paper also tackles the thorny issues of AI-generated text summaries and the cognitive effects of AI dependency. We wrap up by exploring what's really at stake when organisations get the human-AI balance wrong.

Transparency Note:

This paper was written through a collaborative process between human researchers and Claude-Sonnet 4, an AI assistant created by Anthropic. The human author provided the conceptual framework, research direction, source materials, and editorial oversight, while Claude assisted with development, organisation, and contributed to writing. All factual claims, analysis, and conclusions reflect human judgment and verification. This collaborative approach exemplifies the human-centric AI methodology advocated throughout this paper—leveraging AI capabilities to enhance human research and writing while maintaining human authority over content, accuracy, and final decisions. The author believes this transparency aligns with the paper's core argument that effective AI implementation requires clear acknowledgment of both human and machine contributions to collaborative work.

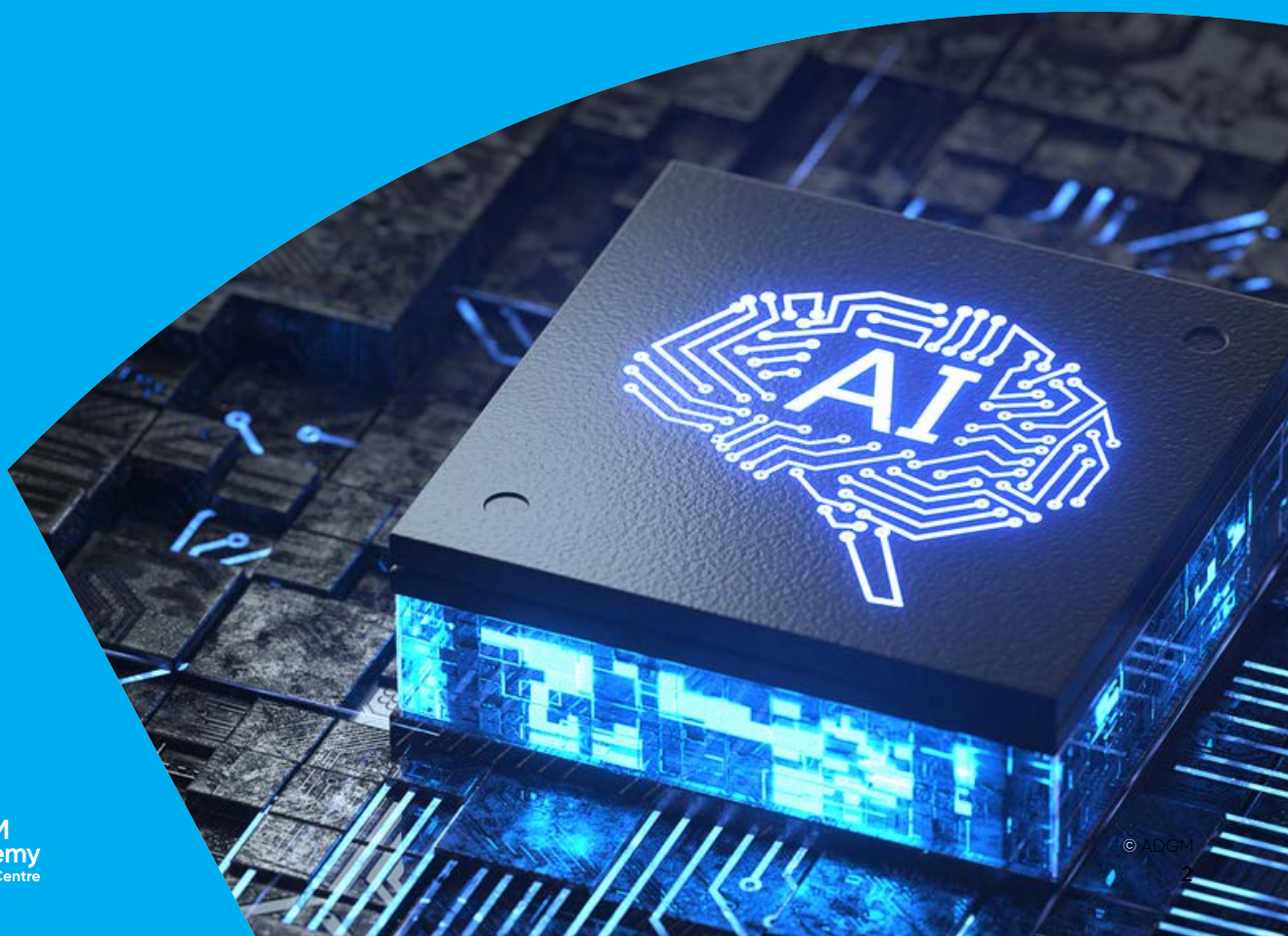


Table of Contents

INTRODUCTION: THE CRITICAL CHOICE

We're living through a moment when artificial intelligence is reshaping entire industries faster than most of us can keep up. The efficiency gains are real, and the opportunities seem endless. But here's the thing that keeps business leaders up at night: should AI tools be calling the shots, or should they stay in their lane as powerful assistants that make humans more capable?

This isn't just a theoretical debate anymore. The stakes have gotten personal, as Joanna Maciejewska captured perfectly in a viral post that struck a nerve with millions: "I want AI to do my laundry and dishes so I can do art and writing, not for AI to do my art and writing so that I can do my laundry and dishes."[1] That sentiment hit home because it gets to the heart of what we're all worried about—are we building AI that frees us up for the work that matters, or are we accidentally creating systems that push us toward the boring stuff while machines handle the creative, meaningful parts of our jobs?

The choice between human-centric and tech-centric AI isn't academic. It's determining right now how AI gets woven into businesses, government agencies, and pretty much every corner of society. Companies are scrambling to adopt AI because competitors are doing it, investors are demanding it, and vendors are promising the moon in terms of efficiency gains. What we're seeing in these early days is a tale of two approaches: some organisations are crashing and burning spectacularly (think lawyers getting sanctioned for fake case citations), while others are discovering that thoughtful human-AI collaboration can produce results neither humans nor machines could achieve alone.

The organisations diving into AI without proper human oversight are walking into minefields. They're facing regulatory fines, angry customers, and—perhaps most damaging—the slow erosion of the human capabilities that drive long-term success. Meanwhile, companies that figure out how to blend human judgment with AI capabilities are gaining advantages that go way beyond simple automation. They're making better decisions, managing risks more effectively, and earning trust from stakeholders who might otherwise be sceptical of AI.

This paper contends: AI delivers its biggest benefits when it's designed to make humans more capable, not when it's built to replace them. The goal should be keeping humans firmly in the driver's seat while leveraging AI's

computational superpowers. This human-centric approach recognises something fundamental—while machines can crunch numbers and spot patterns with exceptional efficiency, humans bring creativity, empathy, ethical reasoning, and strategic vision that no algorithm can match.

THE CENTAUR ADVANTAGE: LESSONS FROM CHESS AND MILITARY STRATEGY

Garry Kasparov's Revolutionary Insights

If you want to understand how humans and AI can work together effectively, start with Garry Kasparov's chess story. This is the guy who famously lost to IBM's Deep Blue back in 1997—a moment that seemed to signal machines were taking over intellectual work. But instead of throwing in the towel, Kasparov did something brilliant. He started experimenting with "centaur chess," where human players team up with AI systems.[2]

What he discovered was game-changing, literally. These human-machine teams didn't just compete with solo players or standalone AI—they dominated them. And this wasn't some temporary fluke. It represented a fundamental shift in what becomes possible when human intelligence meshes thoughtfully with machine capabilities.

The secret sauce comes from how humans and machines complement each other's strengths:

- AI brings computational horsepower, pattern recognition that can process massive datasets without getting tired, and consistent performance regardless of whether it's Monday morning or Friday afternoon
- Humans contribute creativity, gut instincts, strategic thinking, adaptability, and the kind of contextual judgment that only comes from years of real-world experience. Plus, they can navigate ethical dilemmas and understand stakeholder concerns in ways no algorithm can touch

Successful collaboration means humans need to understand both what AI can do brilliantly and where it falls flat. The chess players who learned when to trust AI recommendations and when to override them consistently outperformed those who either ignored the AI completely or followed it blindly.

Military Applications: The Digital Centaur Concept

The U.S. military took Kasparov's insights and applied them in situations where the stakes couldn't be higher. Their "digital centaur" approach strategically divides responsibilities between humans and machines, letting each focus on what they do best.[3] In military contexts, this means machines handle the computational heavy lifting while humans retain control over strategy, ethics, and the final decisions that can mean life or death.

In practice AI systems scan through millions of data points from satellite imagery, communications intercepts, and open-source intelligence at scales that would overwhelm any human analyst. They spot patterns, flag potential threats, and maintain constant surveillance across multiple channels simultaneously.

But—and this is crucial—humans stay in charge of what matters most. They interpret AI-generated intelligence within broader strategic contexts that machines simply can't grasp. They apply ethical frameworks to targeting decisions, weigh long-term political implications, and accept ultimate responsibility for outcomes in ways that no algorithm ever could or should.

One illustrative example comes from intelligence analysis, where AI systems can process satellite imagery, communications data, and open-source intelligence at scales impossible for human analysts. However, the decision to act on identified threats remains firmly in human hands, ensuring accountability and ethical oversight.

THE HIGH COST OF AI OVER-RELIANCE:
REAL-WORLD CAUTIONARY TALES

Recent years have witnessed several high-profile cases where excessive reliance on AI systems led to significant professional and financial consequences, particularly in the legal sector. These cases serve as stark reminders of why human oversight remains essential, especially in high-stakes professional environments.

Legal System Failures

The legal profession has been hit with some astonishing examples of AI over-reliance that resulted in professional sanctions and serious reputational damage. These weren't cases of rogue employees or under-resourced firms—we're talking about experienced professionals at reputable practices who simply trusted AI outputs without doing their homework.

The Vancouver Case: In British Columbia, an attorney found themselves at the centre of Canada's first major AI-in-court fiasco. The attorney was representing a wealthy father in a messy custody dispute, helping him seek permission to take his children to China for a visit. When they asked ChatGPT for relevant case law, the system provided three cases that looked perfect for the client's situation.

Two of those cases would have provided compelling precedent for international travel with children—if they had existed. When opposing counsel couldn't track down the cited cases despite repeated requests, the whole thing unravelled. In his February 2024 ruling, the presiding judge calling the incident "alarming" despite finding no intent to deceive.[4]

The judge's observation was unambiguous: "As this case has unfortunately made clear, generative AI is still no substitute for the professional expertise that the justice system requires of lawyers." The attorney got stuck personally compensating the opposing legal team for the time they spent trying to track down the AI-generated citations. What makes this case particularly eye-opening is that the attorney wasn't some tech novice - they have a PhD in law and five years of practice experience. Yet they fell into the same trap that's catching professionals across industries: assuming that sophisticated technology produces reliable outputs.

The New York Sanctions: The most expensive lesson came from federal court in Manhattan, where attorneys learned that AI mistakes carry very real consequences. Their case involved an employment injury claim against an airline—routine personal injury litigation that should have been straightforward. The legal team used ChatGPT for legal research, something they never tried before. The system delivered several cases involving aviation accidents that seemed directly relevant to their case. The issue? Cases like "Martinez v. Delta Air Lines," "Zicherman v. Korean Air Lines," and "Varghese v. China Southern Airlines" were completely made up.

The judge didn't just fine the attorneys $5,000 each—they delivered a withering rebuke that's become required reading in legal circles. The judge found that the attorneys acted in bad faith, making "acts of conscious avoidance and false and misleading statements to the court."[5] What's particularly troubling is that the attorney actually asked ChatGPT whether the cases were real. The system confidently assured him they existed and could be found in reputable legal databases. This false confidence turned out to be more dangerous than obvious uncertainty would have been.

AI-Driven Biometric Fraud in Africa's Financial Sector

A new trend in Africa demonstrates a new kind of AI-related threat. In sub-Saharan Africa, criminal groups are using AI to conduct biometric fraud on an increasing scale. Biometrics and digital identity initiatives, including remote account opening, are increasingly being put in place across Africa and the Middle East (AME). However, criminals are now actively using generative AI to unlock biometrics security in attacks. As noted by a Smile ID report, biometric fraudsters are creating biometrics deepfakes that replicate fingerprints, facial images, and even voices. Using generative AI, criminals are even using these tools to create synthetic identities to spoof identity biometrics and liveness detection technology which banks and fintech companies were already putting in place to protect accounts and digital transactions. This has led to a sharp increase in financial crime and illicit activity in fraudulent accounts and has made achieving safe digital financial inclusion more difficult. As African cybersecurity firms are now raising the alarm, these deepfakes are now more accurate than before and are being used to target Africa's growing digital economy, a focus that banks and fintechs in emerging markets have increased their efforts. The relative lack of institutional governance and experience with technology in Africa, for example in staff training, is a contributing factor.

The erosion of biometric security with the rise of AI-spoofed fake biometrics calls attention to what can happen when security models are too heavily based on automation and without rigorous independent manual oversight, similar to the limitations with some LegalTech or RegTech cases. A key question is whether AI and its supporting infrastructure are being appropriately configured to support humans. Institutions and governments need to remember that humans are an essential component in customer-facing and internal due diligence security and customer support use cases. AI-based security lures or prompts need adequate controls and checks, such as additional and more intrusive background checks to counter potential spoofing and AI use, for example, to generate false credentials or biometric identifiers. In many ways, the learning curve for AI, both in institutions and cybersecurity environments, is working in reverse and is ripe with opportunities for new and emerging types of financial crime.

The Pattern Behind the Failures: These issues reveal troubling patterns that extend way beyond the courtroom. First, professionals accepted AI outputs without independent verification, showing fundamental gaps in their quality control processes. Second, users failed to recognise AI's tendency to generate plausible but completely false information—a phenomenon experts call "hallucination." Third, many users lacked proper training on AI limitations and verification protocols, despite being experts in their own fields. Finally, these failures happened

even though professional obligations to verify information already existed, suggesting that traditional accountability structures might not be enough for AI-assisted work. The takeaway for business leaders is crystal clear: being competent in one domain doesn't automatically make you competent with AI. Organisations need new training programmes, verification protocols, and accountability structures designed specifically for AI-assisted work.

Financial Services Discrimination

The financial services sector has run into its own set of serious problems with AI bias, particularly in credit decisions and lending. Research consistently shows that AI systems can perpetuate and amplify existing discrimination patterns. A recent prominent credit controversy exemplifies these broader concerns more concretely.

The bank's algorithm for the Apple Card drew fierce criticism for allegedly discriminating against female applicants. Women reported receiving lower credit limits or being denied cards entirely while their husbands with similar financial profiles sailed through. The incident highlighted not just potential gender bias but the critical need for ongoing human oversight of AI decision-making systems.

This wasn't simply a case of biased programmers or insufficient data. Research reveals that predictive tools used for loan approvals are systematically less accurate for minority applicants, but not primarily due to intentional discrimination. Studies of real mortgage data found that differences in approval rates between demographic groups stem partly from minority and low-income populations having less comprehensive credit histories—a data limitation that AI systems can't overcome and may make worse.

The implications are staggering. AI systems trained on historical data will keep perpetuating historical inequalities unless humans actively step in to address these patterns. This requires ongoing human oversight that goes beyond surface-level monitoring to include systematic bias testing and algorithmic auditing. The US-based Consumer Financial Protection Bureau has responded by issuing guidance requiring that even complex AI-driven credit decisions must provide accurate and specific reasons for adverse actions.[6]

Even more concerning, some financial institutions have reduced their human risk management teams as they've ramped up their reliance on automated systems. This creates dangerous single points of failure where AI errors can cascade through entire organisations before anyone notices the problems.

THE CHALLENGE OF TEXT SUMMARISATION: WHEN AI GETS IT WRONG

Here's a problem that's flying under the radar but affecting knowledge workers everywhere: AI text summarisation. People are increasingly relying on AI to summarise lengthy documents—regulations, contracts, research papers—and assuming these summaries accurately capture the important stuff. That assumption can be dangerous, because the process by which AI decides what to include or exclude from summaries remains opaque and prone to significant errors.

The Summarisation Problem

Text summarisation is particularly tricky for AI systems because it requires more than just language understanding—it needs the ability to identify what information is most important within a specific context. These Large Language Models manage summarisation in a fundamentally different way than human beings do. Yes, they can produce shorter versions of texts, but the summaries often lack the contextual anchors that human readers bring to the task. Without knowing the reader's specific role, responsibilities, and decision-making needs, AI systems essentially "summarise to average"—producing generic summaries that may miss details crucial to specific readers. As a recent Apple Research Paper noted in "The Illusion of Thinking", while these programmes may outperform standard models on moderately complex tasks, they appear to be failing entirely on high-complexity problems due to their inherent limitations in reasoning scalability and consistency.

This proves particularly dangerous because evaluating summarisation quality is inherently subjective. There's no single "correct" summary of any complex document. Quality depends entirely on the reader's context, expertise, and intended use of the information. Traditional evaluation metrics focus on surface-level features like length reduction and keyword preservation rather than whether the summary serves the reader's needs.

For regulatory documents and compliance materials, these limitations create particularly high stakes. When knowledge workers assume AI-generated summaries capture all relevant requirements, they may miss critical

details, misinterpret compliance obligations, or fail to identify important exceptions that could expose their organisations to regulatory violations.

The opacity of AI summarisation makes this problem worse. There's no transparent methodology for understanding why certain information was prioritised or excluded, making it impossible to assess whether critical details were overlooked. Unlike human summarisers who can explain their reasoning, AI systems provide no insight into their decision-making process.

THE COGNITIVE IMPACT:
HOW AI USE AFFECTS HUMAN THINKING

Here's something that should worry every business leader: mounting research suggests that our increasing reliance on AI tools may be subtly reshaping how humans think, potentially with long-term implications for critical thinking and intellectual development.

Research on Cognitive Changes

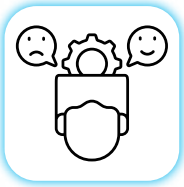
A comprehensive study of 319 knowledge workers conducted by researchers at Microsoft and Carnegie Mellon University revealed a troubling pattern: higher confidence in AI correlates with reduced critical thinking, while higher self-confidence promotes more critical thinking.[7] This suggests that as employees trust AI more, they think for themselves less.

The study found that knowledge workers engage in critical thinking when using AI primarily to ensure work quality—setting clear goals, refining prompts, and assessing AI-generated content against specific criteria. However, several barriers consistently inhibit this reflective process: lack of awareness about AI limitations, time pressures that discourage careful analysis, and difficulty improving AI responses in unfamiliar domains.

Additional research found a significant negative correlation between frequent AI tool usage and critical thinking abilities, mediated by increased cognitive offloading. Studies revealed that younger participants exhibited higher

dependence on AI tools and lower critical thinking scores compared to older participants, suggesting that digital natives may be particularly susceptible to these effects.

This concern has become more urgent with policies such as the recent UAE decision to roll out an AI curriculum across all government schools, from K-12. While equipping students with cutting-edge skills is vital, such early and sustained exposure to AI increases the imperative to safeguard against possible negative impacts on independent reasoning and critical thinking—core abilities that are still forming in children and young adolescents.



The Cognitive Offloading Phenomenon

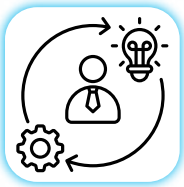
Researchers have identified what they call "Artificial Intelligence Chatbot-Induced Cognitive Atrophy" (AICICA)—a phenomenon describing the potential cognitive consequences of excessive AI reliance.[8]

This concept suggests that over-dependence on AI systems may disproportionately affect younger workers who prioritise convenient access to information over deep learning and critical thinking development.

Cognitive offloading—delegating mental tasks to external systems—isn't inherently problematic. Humans have always used tools to extend our cognitive capabilities, from written language to calculators. However, recent research suggests that the speed and convenience of AI systems may be creating cognitive atrophy characterised by several concerning trends:

- **Reduced Cognitive Effort:** When AI provides immediate answers, users may skip the mental work of analysis, synthesis, and critical evaluation that builds cognitive strength over time.
- **Overconfidence in AI Outputs:** Research shows that AI systems often present information with apparent confidence, lacking explicit representations of uncertainty, which can lead users to accept outputs without adequate scrutiny.
- **Diminished Tolerance for Cognitive Dissonance:** The immediate, synthesised responses from AI may reduce the productive confusion and questioning that typically drives deeper learning and insight.

Studies demonstrate that over-reliance on AI dialogue systems negatively affects critical cognitive capabilities including decision-making, critical thinking, and analytical reasoning. Individuals increasingly favour fast solutions over thoughtful analysis, potentially undermining the cognitive skills that drive innovation and adaptation.

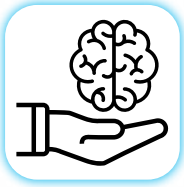


Implications for Professional Development

These cognitive changes have particular implications for professional environments where critical thinking and analytical skills are essential. When employees routinely use AI for tasks that previously required careful analysis—such as reviewing contracts, analysing market data, or synthesising research—they may gradually lose the cognitive muscles that made them effective in these roles.

Studies of human cognition reveal that human thinking is less strictly logical and more grounded in intuitive knowledge that comes from experience and practice. When AI systems handle the analytical heavy lifting, professionals may lose opportunities to develop this intuitive expertise.

However, the research also suggests reasons for optimism. Work on "protective factors" indicates that metacognitive prompts and awareness training can help users maintain critical thinking skills even while using AI tools. Organisations that implement training programmes focused on AI literacy and critical evaluation of AI outputs may be able to capture AI's benefits while preserving human cognitive capabilities.



Protecting Human Intelligence

The solution isn't to avoid AI but to use it more thoughtfully. At the beginning of learning new skills, cognitive effort and persistence remain essential for deep understanding, even when AI assistance is available. During early learning stages, AI use should remain minimal to ensure employees develop fundamental competencies before relying on technological assistance.

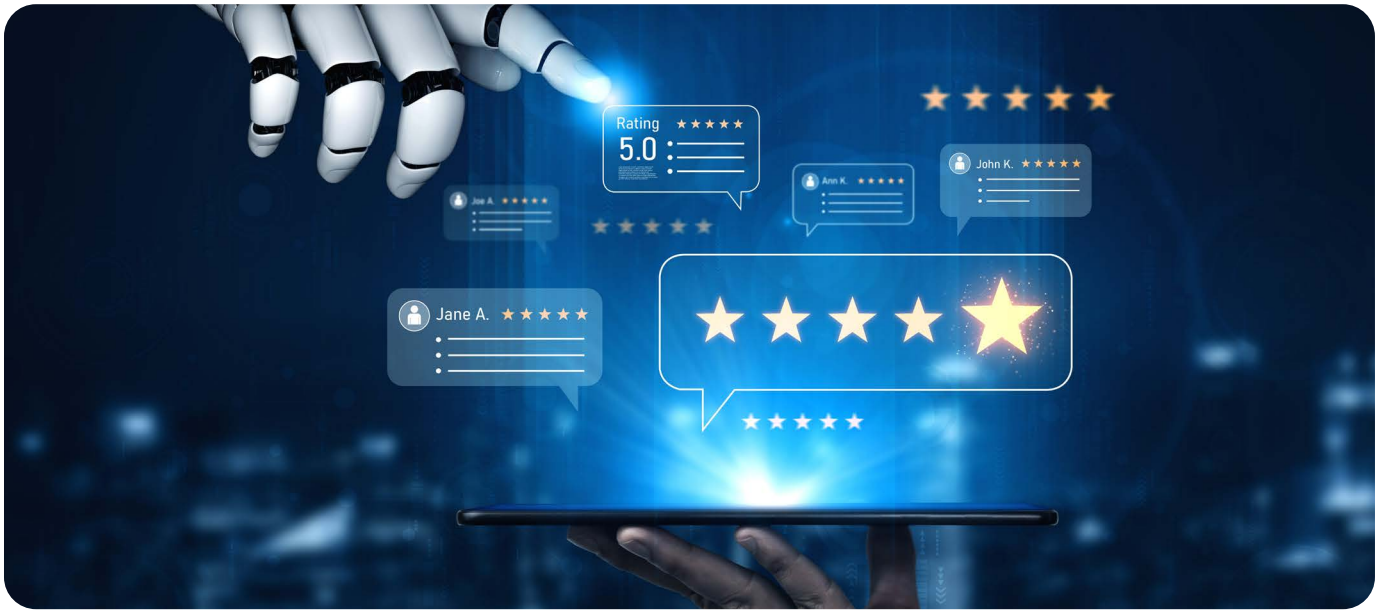
The proposed UAE curriculum initiative provides an opportunity to embed safeguards directly into educational practice at the earliest stages and where it matters most. Courses will need to address the psychology of learning, the risks of automation, and the value of critical thinking to future-proof students' intellectual development. Moreover, teacher professional development should similarly focus on guiding students to use AI as a tool for exploration and evaluation—rather than as a replacement for their own judgment.

Organisations should maintain opportunities for employees to engage in effortful mental activity, ensuring that not all challenging work gets automated away before humans have chances to develop crucial skills. Metacognitive training helps workers become more aware of their thinking processes when using AI, enabling them to recognise when they're deferring too heavily to technological solutions.

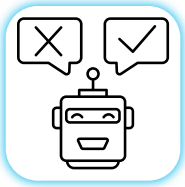
Rather than implementing AI tools comprehensively across all functions simultaneously, graduated integration allows employees to adapt while maintaining critical thinking skills. Critical evaluation protocols should make human analysis mandatory rather than optional, ensuring that AI outputs receive genuine scrutiny rather than perfunctory approval.

Most importantly, organisations should preserve spaces for purely human analysis and decision-making, ensuring that employees continue exercising cognitive capabilities that AI cannot replace and that remain essential for strategic thinking and innovation.

BUILDING AI THAT SERVES HUMANS:
DESIGN PRINCIPLES AND IMPLEMENTATION



To ensure AI systems empower rather than diminish human capabilities, organisations should adopt specific design principles and implementation strategies. Human-centric AI depends on three fundamental design principles: Transparency, Collaboration, and Flexibility.



Transparency: The Foundation of Trust

Creating AI systems that truly serve human needs starts with transparency—not just technical transparency for engineers, but practical transparency that helps users understand how systems reach their conclusions. When loan officers use AI-powered approval systems, they need clear explanations of decision factors that they can communicate to both colleagues and customers.

Transparency is paramount in building trust and enabling effective oversight of AI systems. To ensure AI remains a tool that augments rather than dictates, users must have a clear understanding of how AI-driven decisions and recommendations are generated. This necessitates moving beyond the "black box" approach, where AI outputs are opaque and inscrutable.

Practical applications of transparency include explaining AI recommendations in plain language. For instance, an AI-powered loan application system should not simply deny an application; it should provide clear, concise reasons for the decision, highlighting the key factors that influenced the outcome, such as credit history, income level, or debt-to-income ratio.

Another crucial aspect is visualising data relationships that influence AI outputs. In a sales forecasting tool, for example, the AI might identify correlations between sales performance and various market variables. Presenting these relationships visually, through charts and graphs, allows sales managers to understand which factors the AI considers most important and how they interact to generate predictions.

Documentation of limitations and confidence levels proves equally important. No AI system is infallible, and users

must be aware of potential inaccuracies or uncertainties. A medical diagnosis AI, for instance, should provide not only a list of possible diagnoses but also a confidence score for each, reflecting the level of certainty in its predictions. This empowers doctors to weigh the AI's input against their own clinical judgment.

The goal isn't to overwhelm users with technical details but to provide enough insight for them to make informed decisions about when to trust AI recommendations and when to apply their own judgment.

Collaboration: Humans and AI as Partners

Successful AI implementation promotes genuine teamwork between humans and technology rather than simple task automation. This collaborative approach recognises that optimal outcomes emerge when human and artificial intelligence complement each other's strengths rather than competing for the same roles.

Effective AI system design promotes teamwork between humans and AI, fostering a collaborative environment where machines augment human roles without entirely taking over. This "centaur" approach, inspired by chess and military strategy, recognises the unique strengths of both humans and AI, emphasising that the best outcomes arise when they work in synergy.

To achieve this collaboration, intuitive interfaces are essential. AI tools should be designed to be user-friendly, allowing for seamless interaction and easy input of information. A customer service system that feels natural and helpful can be achieved through well-designed user interfaces that facilitate smooth human-AI interaction.

Furthermore, there must be a clear delineation of AI and human responsibilities. In hiring systems, AI might screen resumes and identify qualified candidates based on objective criteria like education requirements and years of experience. But human hiring managers should retain final decision-making authority, particularly for assessments of cultural fit, communication skills, and leadership potential—qualities that require nuanced judgment about team dynamics and organisational culture.

Override mechanisms prove essential for maintaining human agency when AI gets things wrong. Financial trading platforms should allow experienced traders to deviate from AI investment recommendations based on market insights, geopolitical considerations, or risk factors that may not be captured in algorithmic models. These override capabilities shouldn't be buried in complex menus but should be easily accessible when human expertise suggests alternative approaches.

Clear role delineation prevents confusion about decision-making authority while ensuring accountability remains appropriately assigned to human decision-makers who understand the broader implications of their choices.

Flexibility: Adapting to Human Diversity

Effective AI systems accommodate users with different needs, preferences, and expertise levels rather than forcing everyone to work the same way. This adaptability ensures that AI tools remain valuable across varied organisational contexts and user capabilities.

Effective AI systems aren't one-size-fits-all; they adapt to diverse contexts and cater to users with different needs, preferences, and skill levels. This adaptability is essential for ensuring that AI tools are accessible, user-friendly, and effective across a wide range of situations.

Customisable interfaces allow users to tailor AI systems to their specific requirements rather than accepting one-size-fits-all solutions. Data analytics tools should enable users to choose preferred visualisations, select relevant metrics based on their responsibilities, and configure notification systems that align with their decision-making schedules.

Expertise-based automation levels recognise that novice and expert users benefit from different levels of AI assistance. Beginners often benefit from highly automated systems with extensive guidance, while experts typically prefer advanced tools that offer greater control and flexibility. Design software might offer templates and automated suggestions for newcomers while providing sophisticated customisation capabilities for experienced designers who want to push creative boundaries.

Context-sensitive assistance enables AI systems to recognise situational factors and provide relevant support based on current circumstances. Manufacturing AI should provide different instructions for routine maintenance versus emergency repairs, adapting communication style and urgency level to match the situation and potential consequences.

ALIGNING HUMAN-CENTRIC AI PRINCIPLES WITH POLICY AND PRACTICE

National and Global regulatory strategies are coming into place to set the standards and guidance for responsible and trustworthy AI. For example, the EU AI Act, the first worldwide risk-based regulatory framework for AI, explicitly sets requirements for transparency, human agency/guardrails and accountability for high-risk applications. This new law will set the minimum standard on which to base technology and therefore requires that the principles of Human-Centric AI be considered as the core foundation for AI development. Singapore's Financial Sector AI Explainability Initiative, which is aligned with the Singapore government's overarching AI Principles (FEAT: Fairness, Ethics, Accountability, and Transparency) provides financial institutions with a set of guidelines to ensure transparent, fair, and responsible AI, including those that are human-in-the-loop or human-agent augmentation, to enable business with clear auditing and explainability. The UAE announced its new 10-year National AI Strategy

2031, reaffirming the need to develop AI innovations that are "always in service to humanity" and that it will prioritise ethical AI and public trust; a clear revision to the initial 2021 policy announcement that was notable for omitting oversight of the technology.

FUTURE CONSIDERATIONS: EMERGING AI TRENDS

Contextual AI: Intelligence That Adapts

The first major trend involves contextual AI systems that understand real-world environments and adapt their responses based on user behaviour, preferences, and situational factors in ways that current systems simply cannot match. These systems represent a significant leap beyond today's AI tools, which typically provide the same responses regardless of context.

Contextual AI excels at personalising experiences across diverse users and situations, recognising that what works for one person in one context may be completely inappropriate for another person in different circumstances. These systems integrate information from multiple sources to provide relevant support that feels natural and helpful rather than generic and robotic. Perhaps most importantly, they learn from interactions to improve future performance, becoming more valuable over time rather than remaining static.

In healthcare settings, contextual AI can tailor treatment recommendations based on a patient's complete medical history, genetic factors, lifestyle choices, and personal preferences. But the critical distinction lies in implementation: these systems support human clinicians by providing comprehensive, personalised information while leaving final treatment decisions firmly in human hands. The AI enhances the doctor's ability to consider multiple factors simultaneously, but the doctor retains full responsibility for patient care.

Educational applications demonstrate similar potential. AI can adapt learning materials to students' progress rates, learning styles, and individual challenges while supporting human teachers rather than replacing them. The technology enables more personalised education at scale, but teachers remain essential for motivation, emotional support, and the complex judgment calls that effective education requires.

Agentic AI: Autonomy Within Boundaries

Agentic AI represents an even more significant development—systems that demonstrate greater autonomy and can adapt dynamically to changing circumstances without explicit programming for every possible scenario. These sophisticated systems can take initiative within carefully defined parameters, learn from past experiences to improve future actions, and manage complex tasks with minimal direct human intervention.

The efficiency potential proves enormous. Agentic AI could revolutionise supply chain management, financial planning, customer service, and countless other business processes by handling routine decisions automatically while escalating unusual situations to human oversight. These systems could operate 24/7 without fatigue, processing information and making decisions at speeds impossible for human operators.

But the risks are equally significant. Organisations will need robust ethical oversight and clear accountability mechanisms to ensure these autonomous systems operate within appropriate boundaries while maintaining human control over critical decisions. The challenge lies in defining those boundaries precisely enough to prevent harmful actions while allowing enough flexibility for the systems to provide genuine value.

Ethical Implications of Advanced AI

The development of increasingly autonomous AI systems raises profound ethical considerations that organisations cannot afford to ignore. Transparency becomes even more critical as systems become more sophisticated users must understand how AI systems make recommendations or decisions, even when those systems operate with greater independence.

Accountability structures must evolve to handle situations where AI systems make decisions autonomously within predefined parameters. Clear lines of responsibility must exist for AI-assisted outcomes, but organisations must also develop new frameworks for situations where AI systems act independently within authorised bounds.

Trust represents perhaps the greatest challenge. Systems must be designed to maintain user confidence through reliable performance and clear communication about their capabilities and limitations. Without sufficient transparency, the deployment of complex AI systems risks eroding trust in ways that could undermine the entire AI enterprise.

Consider financial institutions implementing AI-driven loan approval processes. If these systems operate without explaining rejection rationales, customers may perceive bias or unfairness, damaging organisational reputation and potentially reinforcing existing societal inequities. The solution lies not in avoiding advanced AI but in implementing it with appropriate transparency and human oversight.

THE STAKES OF GETTING IT RIGHT....OR WRONG

The consequences of failing to strike the right balance between human-centric and tech-centric AI are profound, affecting not only businesses but society at large. What's at stake in this balance extends far beyond organisational efficiency or technological sophistication—it affects fundamental aspects of human agency, social equity, and institutional trust.

Erosion of Trust and Institutional Credibility

If AI systems are perceived as opaque or untrustworthy, individuals and organisations may resist adoption despite potential benefits. The legal cases documented in this paper illustrate how AI failures can devastate professional reputations and institutional credibility. When lawyers submit briefs containing fictitious cases generated by AI, or when AI systems perpetuate discriminatory lending practices, public trust in these technologies erodes rapidly.

This resistance stems from legitimate concerns about accountability, explainability, and alignment with human values. A loss of trust could significantly limit AI's potential to drive innovation and efficiency across sectors. Trust depends on transparency—not necessarily about algorithmic details, but about how decisions are made, what factors influence outcomes, and where human oversight occurs. Organisations that fail toce and external backlash.

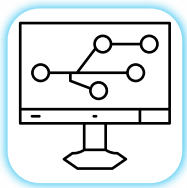
Disempowerment of Individuals

Over-reliance on AI decision-making could lead to what some experts call "mental obesity," where humans become passive recipients of machine-generated decisions rather than active, critical thinkers. The cognitive research discussed in this paper shows concerning trends: frequent AI users demonstrate reduced critical thinking abilities, and younger workers show particular vulnerability to cognitive offloading effects.

This disempowerment risks creating a workforce ill-equipped to address novel challenges or adapt to changing circumstances. The risks extend beyond workforce capabilities to questions of agency and autonomy. When systems make decisions without meaningful human input or oversight, individuals lose control over processes that affect their lives and work. This loss of agency contradicts the fundamental purpose of technology: to empower rather than diminish human potential.

Ethical and Social Inequities

The integration of AI carries significant ethical implications, particularly concerning fairness and equity. A purely tech-centric approach to AI implementation often prioritises efficiency and automation, potentially overlooking or even exacerbating existing societal biases and creating new forms of inequity. The financial services examples in this paper demonstrate how AI systems can perpetuate discrimination, costing minority borrowers hundreds of millions of dollars annually in additional interest payments. AI-led hiring systems, if not designed and monitored with a human-centric approach, might inadvertently reinforce historical patterns of discrimination. These systems could be trained on data that reflects past hiring practices, which may have been skewed against certain demographic groups. As a result, the AI might perpetuate those biases, limiting opportunities for qualified candidates from underrepresented backgrounds. Furthermore, the deployment of AI can create information asymmetries that benefit technologically sophisticated organisations at the expense of others. Companies with greater resources and expertise in AI development and implementation may gain a competitive advantage, while smaller businesses or individuals may be left behind. This can widen the gap between the "haves" and "have-nots" in the digital age.



Economic and Organisational Fragility

Businesses that adopt overly tech-centric AI models expose themselves to several significant risks that can undermine their stability and long-term success. These risks often stem from a failure to adequately consider the human element in AI implementation.

One key risk is employee disengagement. When workers perceive that AI is being used to replace them rather than augment their capabilities, it can lead to feelings of insecurity, resentment, and a decline in morale. This disengagement can negatively impact productivity, innovation, and employee retention.

Another significant risk is customer backlash. The legal cases examined show how AI failures can result in substantial financial penalties and reputational damage. When a law firm faces sanctions for AI-generated fake citations, or when financial institutions must pay millions in discrimination settlements, the costs extend far beyond the immediate penalties.

Organisations that remove human judgment from critical processes become more vulnerable to systemic failures. Over-reliance on AI without adequate human oversight can create brittle systems that are unable to adapt to unexpected challenges or recover quickly from disruptions. In contrast, organisations that implement human-AI collaboration models preserve human oversight while leveraging technological capabilities, creating systems that are more adaptable and resilient.

Missed Opportunities for Innovation

Perhaps the most significant consequence of a tech-centric approach is that organisations miss the transformative potential of true human-AI collaboration. When AI is viewed primarily as a tool to replace human workers, it severely limits the possibilities for innovation. True innovation emerges at the intersection of human creativity, strategic thinking, and ethical considerations

with machine intelligence, where each enhances the other's capabilities.

Human-AI collaboration can lead to the discovery of novel applications that neither humans nor machines would identify independently. AI can handle complex calculations, data analysis, and repetitive tasks, freeing up human employees to focus on the higher-level creative and strategic aspects of problem-solving. This combination can lead to more effective and innovative solutions that create value extending far beyond simple efficiency gains.

CONCLUSION: THE PATH FORWARD

To avoid the pitfalls of tech-centric AI and fully realise the transformative potential of artificial intelligence, organisations must decisively embrace a human-centric approach to AI development and implementation. This fundamental shift involves several key principles that collectively ensure AI enhances rather than diminishes human capabilities and agency.

First, it requires designing AI systems that empower individuals in the workplace, rather than diminishing their roles or replacing them entirely. This means focusing on how AI can augment human capabilities, automate mundane tasks, and free up employees to focus on more strategic, creative, and fulfilling work. The digital centaur model provides a proven framework for organisations seeking to leverage AI's capabilities while preserving and enhancing human agency and expertise.

Second, it prioritises transparency in how AI generates recommendations or decisions. The legal failures documented in this paper could have been avoided with proper verification processes and clear understanding of AI limitations. This transparency is crucial for building trust and enabling human users to understand, validate, and, if necessary, override AI outputs.

Third, it fosters genuine collaboration between humans and machines, consciously leveraging their complementary strengths. This "centaur" approach recognises that humans bring critical thinking, creativity, ethical judgment, and contextual awareness, while AI excels at data analysis, pattern recognition, and computational tasks.

Fourth, it builds in robust ethical safeguards that reflect both organisational values and broader societal expectations. This includes addressing potential biases in AI algorithms, ensuring fairness and equity, protecting privacy, and establishing clear accountability for AI-driven actions. The financial services discrimination examples show the real-world costs of failing to implement these safeguards.

Fifth, it creates clear governance structures that ensure appropriate oversight and accountability throughout the AI lifecycle. This involves defining roles and responsibilities, establishing audit trails, and implementing mechanisms for monitoring AI performance and addressing any unintended consequences. The text summarisation challenges highlighted in this paper underscore the need for human verification processes, particularly for critical documents.

Organisations must also address the cognitive impacts of AI use by implementing training programmes that maintain and strengthen human critical thinking skills. The research on cognitive offloading suggests that with proper awareness and metacognitive training, professionals can harness AI's benefits while preserving their analytical capabilities.

The path forward requires deliberate choices about AI implementation philosophy rather than simply adopting available technologies because competitors use them or vendors promise efficiency gains. Leaders must focus systematically on how AI can amplify human capabilities while maintaining human authority over critical decisions and preserving the cognitive skills essential for innovation and adaptation.

This approach requires investment in training, governance structures, and oversight mechanisms, but produces sustainable benefits that justify these costs. Organisations mastering this balance will thrive in the AI-powered future, while those failing to maintain human-centricity risk being left behind by competitors who understand that the most sophisticated technology enhances rather than replaces human potential.

Embracing this human-centric philosophy and viewing AI as an enabler rather than a replacement allows organisations to unlock AI's full potential to drive innovation, enhance decision-making, and ultimately improve lives. The future of AI is not predetermined; rather, it will be shaped significantly by the choices we make today regarding how these powerful tools are designed, deployed, and governed.

Ultimately, ensuring that AI remains human-led represents not just a technological imperative for businesses but also a vital social and ethical one. By prioritising human agency, creativity, and judgment alongside technological advancement, we can create a future where digital tools serve humanity's highest aspirations, expand our potential and reinforce our autonomy, instead of narrowing our capabilities or diminishing our control.

In navigating the AI frontier, the fundamental question isn't whether to adopt AI—it's whether to use it to build a future that augments human capabilities or one that diminishes them. The choice defines not just organisational success, but the nature of work and human agency in the digital age. The "digital centaur"—symbolising the powerful partnership of human wisdom and machine intelligence—provides a compelling and inspiring vision for this collaborative and human-centred future.

REFERENCES

1. Joanna Maciejewska (@AuthorJMac), post on X (formerly Twitter), March 29, 2024. The post went viral with nearly 3 million views, resonating widely with artists and knowledge workers concerned about AI's direction. See: https://scoop.upworthy.com/author-explains-why-ai-should-do-mundane-jobs-instead-of-art-and-perfectly-hits-the-nail-on-the-head

2. Garry Kasparov, Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins (PublicAffairs, 2017).

3. William J. Barry, Chase Metcalf, and Blair Wilcox, "Strategic Centaurs: Harnessing Hybrid Intelligence for the Speed of AI-Enabled War," Modern War Institute at West Point, January 17, 2025, https://mwi.westpoint.edu/strategic-centaurs-harnessing-hybrid-intelligence-for-the-speed-of-ai-enabled-war/.

4. Consumer Financial Protection Bureau, "CFPB Issues Guidance on Credit Denials by Lenders Using Artificial Intelligence," September 19, 2023. Available at: https://www.consumerfinance.gov/about-us/newsroom/cfpb-issues-guidance-on-credit-denials-by-lenders-using-artificial-intelligence/

5. Zhang v. Chen, 2024 BCSC 287, Justice David Masuhara, February 26, 2024. See: CBC News, "B.C. lawyer reprimanded for inserting fake cases invented by ChatGPT into court documents," February 27, 2024. https://www.cbc.ca/news/canada/british-columbia/lawyer-chatgpt-fake-precedent-1.7126393

6. Mata v. Avianca, Inc., No. 22-cv-1461 (PKC), 2023 WL 4114965 (S.D.N.Y. June 22, 2023).

7. Smile ID. (2024). The State of KYC and Fraud in Africa Report. https://usesmileid.com/thank-you/fraud-report-2024?cta=https://bit.ly/fraud-report-pdf

8. Shojaee, P., Mirzadeh, I., Alizadeh, K., Horton, M., Bengio, S., & Farajtabar, M. (2025). The Illusion of Thinking: Understanding the Strengths and Limitations of Reasoning Models via the Lens of Problem Complexity. Apple.

9. Hao-Ping (Hank) Lee, Advait Sarkar, Lev Tankelevitch, Ian Drosos, Sean Rintel, Richard Banks, Nicholas Wilson, "The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers," CHI 2025 Conference Proceedings, Microsoft Research, January 2025. Available at: https://www.microsoft.com/en-us/research/publication/the-impact-of-generative-ai-on-critical-thinking-self-reported-reductions-in-cognitive-effort-and-confidence-effects-from-a-survey-of-knowledge-workers/

10. The National News. (2024, February 27). UAE schools to introduce AI curriculum from KG to grade 12. The National.

11. Dergaa et al., "From tools to threats: A reflection on the impact of artificial-intelligence chatbots on cognitive health," PMC Health Sciences, 11(4), 2023. The research introduces the concept of Artificial Intelligence Chatbot-Induced Cognitive Atrophy (AICICA) and its potential impacts on critical thinking capabilities.

12. European Commission (2024). The AI Act: The EU's Artificial Intelligence Act. Retrieved from https://artificialintelligenceact.eu/#:~:text=The%20AI%20Act%20is%20a,legal%20framework%20for%20artificial%20intelligence.

13. Monetary Authority of Singapore. (2018, November 12). FEAT Principles – Principles to Promote Fairness, Ethics, Accountability & Transparency in the Use of AI and Data Analytics in Singapore's Financial Sector. https://www.mas.gov.sg/~/media/MAS/News%20and%20Publications/Monographs%20and%20Information%20Papers/FEAT%20Principles%20Final.pdf

14. UAE Artificial Intelligence Office. (n.d.). UAE National Strategy for Artificial Intelligence 2031. https://ai.gov.ae/strategy/

ACKNOWLEDGEMENTS

This paper was developed in collaboration between the ADGM Academy Research Centre, Rabdan Academy and Standard Chartered Bank.

ADGM Academy Research Centre

The ADGM Academy Research Centre brings together an ecosystem of academics, financial industry practitioners, government and technology experts to unlock the shared potential to improve the financial environment in MENA and beyond. The financial industry continues to transform at a rapid pace with new technologies, disruptors, threats and opportunities appearing all the time. Independent research is crucial to be able to understand and utilise this transformation for the benefit of your business, your customers and society in general. The Research Centre provides that understanding through insights developed in collaboration with the academic community.

Rabdan Academy

Rabdan Academy (RA) is a government-owned education institution established to coordinate and enhance learning outcomes for organisations and individuals in the Safety, Security, Defence, Emergency Preparedness and Crisis Management (SSDEC) Sectors. The Academy is the first higher education institution in the world specialised in the SSDEC domain that achieves top '5 star' ratings in the 2 categories of Teaching and Employability under the QS Stars University Rating System.

Standard Chartered Bank

Standard Chartered Bank is a leading international banking group, with a presence in 54 of the world's most dynamic markets. Our purpose is to drive commerce and prosperity through our unique diversity, and our heritage and values are expressed in our brand promise, here for good.

Standard Chartered PLC is listed on the London and Hong Kong stock exchanges.

For more stories and expert opinions please visit Insights at sc.com. Follow Standard Chartered on X, LinkedIn, Instagram, and Facebook.

ADGM ACADEMY RESEARCH CENTRE

Innovating Knowledge, Empowering Change

The **ADGM Academy Research Centre**, part of ADGM Academy, unites academics, financial practitioners, government, and technology experts to drive innovation and enhance the financial landscape in the UAE, MENA region, and beyond. As the financial sector evolves with new technologies, disruptors, and opportunities, independent research is vital to harness these changes for the benefit of businesses, customers, and society. Through collaborative insights with the academic community, the Research Centre delivers the expertise needed to navigate and capitalise on this dynamic transformation.

ADGM Academy, the knowledge hub of ADGM, is shaping the future of banking, finance, digital innovation, and public services in the region. Committed to aligning with the UAE's vision for economic leadership, we deliver cutting-edge experiential programmes that empower both graduates and professionals and drive industry growth. As a trailblazer in financial and digital training, we collaborate with top industry experts, leading professional organisations, and renowned academic institutions to create innovative, certified programmes. Join us on a transformative journey where world-class education meets opportunity, paving the way for a stronger, smarter financial industry.

Stay up to date with ADGM Academy Research Centre.

adgmacademy.com research@adgm.com

FOLLOW US
ON OUR SOCIAL NETWORKS

Level 20, Al Maqam Tower, ADGM Square, Al Maryah Island, PO Box 111999 Abu Dhabi, UAE
الطابـق 20، بـرج المقام، مربعة أبوظبي العالمي، جزيرة الماريـه، ص ب 111999، أبوظبي، الإمارات العربية المتحدة

T +971 2 333 8500 adgmacademy.com