Navigating the Shifting Sands:

Human Expertise, AI Augmentation, and the Future of Specialized Professions

Salim Al Barami
Yuri Yoshida
Ruslan Asfandiyarov

Contents

1. Abstract

This report examines the evolving landscape of specialized professions in the face of rapid advancements in Artificial Intelligence (AI), particularly as we approach Artificial General Intelligence (AGI). It explores how human experts can leverage their unique skills and integrate AI to not only remain indispensable but also to enhance their capabilities and preserve human agency in an increasingly AI-driven world. Inspired by concepts of enhanced human capability in the age of AI, this report synthesizes recent research on systemic risks (Uuk et al. 2024), appropriate trust (Mehrotra et al. 2024), collaborative workflows (Puerta-Beldarrain et al. 2025), and high-level governance initiatives (ITU. 2024) to provide a forward-looking analysis. It identifies key challenges, opportunities, and strategies for professionals to adapt and thrive. The report argues that proactive preparation, a focus on uniquely human skills, the cultivation of appropriate trust in AI systems, and the strategic adoption of human-AI collaborative models are crucial for ensuring a future where human expertise is not only preserved but amplified by AI. The report gives special emphasis to the *AI governance paradox* (ITU. 2024) and the threat of *gradual disempowerment* (Kulveit et al. 2025).

2. Introduction

2.1 The AI Revolution in Specialized Professions

The accelerating progress of AI, particularly the rise of sophisticated Large Language Models (LLMs) and generative AI, is fundamentally altering the landscape of specialized professions. AI systems now demonstrate capabilities that rival or surpass human performance in tasks previously considered the exclusive domain of human experts, including medical diagnosis (Vaccaro, Almaatouq, and Malone 2024), legal research (Boussioux et al. 2024), and creative content generation (Puerta-Beldarrain et al. 2025). This progress, with some perspectives highlighting its potential to be as transformative as historical technological shifts (Mayer et al. 2025), is driven by advancements in LLMs and AI technologies. Researchers estimate significant long-term opportunities related to AI-driven productivity growth (Mayer et al. 2025), highlighting the economic importance of understanding and adapting to these changes. However, this revolution, while offering immense potential, also presents significant challenges, including the "*AI governance paradox*" (ITU 2024), where both regulation and the technological tools for oversight struggle to keep pace, and the systemic existential risk of *gradual disempowerment* (Kulveit et al. 2025), where incremental AI advancements could lead to a substantial and potentially irreversible loss of human influence.

2.2 Core Research Question and Scope

This report addresses the core question: How can human experts leverage their unique capabilities and enhance them with the help of AI, ensuring their indispensability and preserving

human agency in an AI-dominated world? We explore six key areas, with a forward-looking perspective:

1. The evolving impact of AI, including systemic risks and gradual disempowerment

2. Human-AI integrated workflows as a pathway to augmented expertise

3. The necessary transformation of expert roles

4. Proactive resilience strategies for professionals

5. Case studies demonstrating successful human-AI collaboration

6. Policy and governance recommendations for a future of empowered expertise

2.3 Strategic Vision and Key Arguments

This report argues that the future of expertise is not about replacement by AI, but about *strategic integration and augmentation*. Some analyzes suggest that a key factor in realizing AI's potential is leadership's role in adoption and scaling (Mayer et al. 2025), underscoring the importance of proactive organizational strategies. By embracing lifelong learning, cultivating uniquely human skills, fostering appropriate trust (Mehrotra et al. 2024) in AI systems, and actively shaping the development and deployment of AI, human experts can ensure their continued relevance, enhance their capabilities, and maintain agency in an AI-driven world. Proactive preparation and adaptation are key to navigating the challenges and maximizing the opportunities presented by AI.

3. The Evolving Impact of AI on Specialized Professions: A Future-Oriented Perspective

3.1 From Automation to Augmentation, Displacement, Systemic Risk, and the Threat of Gradual Disempowerment

AI's evolution transcends simple task automation; increasingly, AI systems execute tasks demanding judgment, expertise, and complex decision-making, traditionally the domain of human professionals (Boussioux et al. 2024; Dell'Acqua et al. 2023; Felin and Holweg. 2024). This shift signifies AI's expanding reach into cognitive functions, fundamentally altering knowledge work (Mayer et al. 2025). Driven by advancements in LLMs, processing power, memory, reasoning, agentic capabilities, multimodality, and hardware (Mayer et al. 2025), AI's transformative potential necessitates a revised perspective. We must move beyond viewing AI merely as a tool and recognize its capacity to reshape professional landscapes and introduce systemic risks (Uuk et al. 2024). Furthermore, the potential for *gradual disempowerment* (Kulveit et al. 2025) warrants attention, where incremental AI progress, even without malicious intent, could subtly erode human influence across economic, cultural, and political systems.

Key themes emerge from the literatures:

- **Task Deconstruction:** AI's impact is granular, affecting specific tasks within professions rather than entire jobs. Professionals will likely see a redistribution of tasks, with AI

handling some and humans focusing on others (Boussioux et al. 2024; Dell'Acqua et al. 2023).

- **Skill Shift:** Demand is *accelerating* for uniquely human skills that complement AI: critical thinking, complex problem-solving, creativity, ethical reasoning, and emotional intelligence. The ability to *contextualize and interpret* AI-generated insights will become a core competency (Felin and Holweg 2024).

- **Vulnerability of Specialized Professions:** High-skill professions, previously considered immune to automation, are now facing potential disruption. AI's ability to perform tasks requiring expertise challenges the traditional role of human experts (Boussioux et al. 2024; Vaccaro, Almaatouq, and Malone 2024).

- **Systemic Risks:** The systemic risks posed by AI advancements have been extensively categorized by Uuk et al. (2024) into 13 distinct risk areas, which provide a structured approach to understanding AI's broader societal impact. These categories include (but are not limited to):

    1. **Control Risks**: The potential for AI systems to operate beyond intended human oversight, leading to *unintended or harmful decision-making***.**

    2. **Democracy Risks**: AI-driven systems influencing *political discourse and public opinion***,** potentially undermining democratic institutions.

    3. **Discrimination Risks**: The reinforcement or amplification of *existing biases* in decision-making processes, affecting marginalized communities.

    4. **Economic Risks**: AI's potential to *disrupt labor markets***,** exacerbate inequality, and *centralize economic power* in the hands of a few entities.

    5. **Environmental Risks**: The *resource-intensive nature* of AI development contributing to environmental degradation.

    6. **Fundamental Rights Risks**: Threats to *privacy, security, and freedom of expression* as AI systems mediate human interactions.

    7. **Governance Risks**: Challenges in *regulating AI* due to its rapid evolution and *cross-border influence*.

    8. **Security Risks**: The use of AI in cyber threats, misinformation campaigns, and *automated warfare*.

- **Gradual Disempowerment:** This represents a significant, long-term threat. Incremental advances, driven by market incentives and competitive pressures, could lead to a gradual erosion of human influence and control across critical societal systems (Kulveit et al. 2025).

The current literatures indicate a consensus on AI's transformative potential, but debates continue regarding the balance between augmentation and displacement, the severity and likelihood of systemic risks, and the *plausibility and timescale* of gradual disempowerment.

3.2 Anticipating Future Impacts on Specific Professions

3.2.1 Overview and Key Considerations

AI is increasingly reshaping professional roles across various sectors, offering both opportunities for efficiency and innovation while presenting challenges related to skill shifts, ethical considerations, and systemic risks. The extent of AI's impact varies depending on the nature of the profession, the level of human-AI integration, and the potential for disempowerment. This section examines how AI is influencing key industries, highlighting both benefits and concerns, and emphasizing the continued need for human oversight and strategic adaptation.

3.2.2 Core Professional Domains

Healthcare and Medicine

- **Medicine:** AI is increasingly being integrated into diagnosis, treatment planning, and drug discovery, with medical professionals playing an active role in refining and overseeing AI-assisted decision-making (Mosqueira-Rey et al. 2023). These advancements enhance accuracy and efficiency, yet they also introduce challenges such as deskilling of medical professionals, over-reliance on automated systems, and algorithmic bias (Mayer et al. 2025). Consequently, the role of human practitioners is shifting toward validating AI-generated insights, prioritizing patient-centered care, and ensuring ethical oversight, maintaining alignment between AI-driven medical decisions and expert judgment while fostering interpretability and trustworthiness in clinical practice (Mosqueira-Rey et al. 2023).

Legal Sector

- **Law:** AI is increasingly used in legal research, case analysis, and compliance automation (ITU. 2024; Pew Research Center. 2023). These applications enhance efficiency in legal services but introduce concerns about fairness, transparency, and potential biases embedded in AI models. The use of AI in decision-support systems for legal rulings necessitates careful oversight to ensure ethical and unbiased outcomes. Legal professionals must focus on interpretation, argumentation, and verifying AI-generated recommendations to prevent unjust or opaque legal conclusions.

Financial Services

- **Finance:** AI powers algorithmic trading, risk assessment, and fraud detection (Vaccaro, Almaatouq, and Malone. 2024; Mitchell et al. 2025). While improving efficiency, AI-driven finance introduces risks such as market manipulation, systemic instability, and market volatility risks (Uuk et al. 2024). The integration of AI in financial markets also raises concerns about automated decision opacity and algorithmic bias, necessitating explainability and fairness mechanisms (Mitchell et al. 2025). Human expertise remains essential for strategic decision-making, risk mitigation, and regulatory compliance to ensure AI-driven markets remain transparent and accountable.

Academic and Research

- **Academia:** AI is increasingly used for literature reviews, hypothesis generation, and academic writing assistance (Mehrotra et al. 2024). While AI enhances research efficiency, it also raises concerns about over-reliance on AI-generated content and maintaining academic integrity (Nguyen et al. 2024). Uuk et al. (2024) discuss systemic

risks of AI applications, including its potential to reshape knowledge creation. However, concerns about deskilling in academia are more closely tied to studies on human-AI collaborative workflows and evolving academic roles. To mitigate these risks, academics should integrate hybrid AI workflows, where AI supports data synthesis, but human researchers maintain control over critical evaluation, theoretical development, and ethical oversight (Nguyen et al. 2024).

Creative and Cultural Industries

- **Creative Industries:** AI tools assist with content creation, personalization, and user interaction (Puerta-Beldarrain et al. 2025; Felin and Holweg. 2024; Boussioux et al. 2024). While this enhances creativity and efficiency, it also raises concerns about displacing human artists, reducing originality, and homogenizing content (Boussioux et al. 2024). Felin and Holweg (2024) highlight that AI-generated outputs, while novel, often lack deep contextual reasoning and human emotional depth, further emphasizing the need for human involvement in creative decision-making. Hybrid AI-human collaboration models should ensure that AI augments rather than replaces human creativity, particularly in fields requiring originality and cultural sensitivity (Nguyen et al. 2024).

3.2.3 Emerging Professional Transformations

Education and Training

- **Education**: AI-driven adaptive learning systems, grading automation, and intelligent tutoring improve efficiency but may lead to reduced teacher involvement in personalized education (Mehrotra et al. 2024; Uuk et al. 2024; Nguyen et al. 2024). There are also risks of biased AI models influencing student learning outcomes, emphasizing the need for ethical AI oversight and transparency in educational applications. Human educators must focus on mentorship, ethical oversight, and critical thinking development to counterbalance AI-driven automation.

Media and Communication

- **Journalism**: AI is increasingly used for automated news generation and fact-checking, raising concerns about bias and misinformation (Uuk et al. 2024; Mitchell et al. 2025). AI-driven content prioritization risks amplifying engagement-driven news over investigative reporting. Human oversight and editorial accountability remain essential to maintaining journalistic integrity. While AI tools improve reporting efficiency, they also risk reinforcing pre-existing biases in news coverage and prioritizing engagement-driven content over investigative depth. Hybrid AI-human editorial workflows should emphasize fact verification, transparency, and journalistic accountability to uphold media integrity in an AI-enhanced news ecosystem. Journalists must prioritize fact verification, in-depth analysis, and investigative reporting to maintain journalistic integrity in an AI-enhanced media landscape.

Technical and Engineering Fields

- **Engineering**: AI enhances predictive maintenance, process optimization, and design simulations, improving efficiency and reducing costs (Vaccaro, Almaatouq, and Malone. 2024; Uuk et al. 2024). Machine learning models anticipate system failures, minimizing downtime and operational risks (Mehrotra et al. 2024; Mosqueira-Rey et al. 2023). While AI-driven automation accelerates engineering workflows, the lack of transparency in AI decision-making raises concerns about reliability and accountability (Nguyen et al.

2024). Engineers must ensure AI applications align with safety regulations and ethical standards through rigorous validation and oversight.

- **Cybersecurity**: AI improves threat detection, intrusion response, and risk assessment, yet adversarial AI techniques create new security challenges (Mitchell et al. 2025). Broader systemic risks, including automated cyber threats and AI-driven misinformation campaigns, further complicate digital security (Uuk et al. 2024). Cybersecurity professionals will need to focus on AI adversarial defense strategies, ethical hacking, and human-led security audits to maintain resilience against evolving AI-driven threats. Cybersecurity professionals will need to focus on AI adversarial defense strategies, ethical hacking, and human-led security audits to maintain resilience against evolving AI-driven threats.

Administrative and Operational Roles

- **Healthcare Administration**: AI is increasingly used to automate administrative tasks in healthcare, such as medical record management, insurance claims processing, and hospital workflow optimization (Mehrotra et al. 2024; Mosqueira-Rey et al. 2023; Costa and Moore. 2025; Putty 2025). While AI enhances efficiency, it also raises concerns regarding data privacy, security vulnerabilities, and ethical challenges in AI decision-making, particularly in automated systems that handle sensitive information (Mitchell et al. 2025; ITU. 2024). Ensuring compliance with regulatory standards and implementing AI auditing mechanisms is essential to safeguarding patient rights and maintaining trust in AI-driven healthcare systems. The increasing reliance on AI in healthcare administration necessitates robust data protection policies, AI auditing mechanisms, and human oversight in critical decision-making processes to safeguard patient rights and ensure regulatory compliance.

- **Supply Chain & Logistics**: AI is increasingly utilized for process automation, demand forecasting, and inventory management, improving operational efficiency. However, over-reliance on AI can lead to supply chain disruptions, reduced adaptability in crisis situations, and increased vulnerability to algorithmic inefficiencies. To mitigate these risks, human oversight, contingency planning, and adaptive decision-making frameworks should be integrated into AI-driven supply chain operations to maintain resilience and responsiveness. However, over-reliance on AI can lead to supply chain disruptions, reduced adaptability in crisis situations, and increased vulnerability to algorithmic inefficiencies. To mitigate these risks, human oversight, contingency planning, and adaptive decision-making frameworks should be integrated into AI-driven supply chain operations to maintain resilience and responsiveness.

- **Customer Service & Retail**: AI-powered chatbots, recommendation engines, and dynamic pricing models are reshaping customer interactions (Felin and Holweg. 2024). While improving efficiency, AI-driven retail may result in depersonalized customer experiences and workforce displacement. Businesses should balance AI automation with human-led customer engagement strategies to preserve personalized service quality.

3.2.4 Future Implications and Adaptations

As AI continues evolving, professionals across industries must adapt by developing AI literacy, critical evaluation skills, and hybrid collaboration strategies. The key to mitigating AI-related

risks lies in proactive policy development, human-AI integration frameworks, and ongoing ethical oversight. This ensures that AI serves as an augmentative tool rather than a disruptive force, preserving human agency and expertise in an AI-driven world.

4. Human-AI Integrated Workflows: A Path to Augmented Expertise and Empowerment

As AI continues to advance, human-AI integration is becoming a critical factor in shaping the future of expertise. AI systems can support human professionals in a wide range of tasks, from decision-making to creative problem-solving. However, the effectiveness of these collaborations depends on the design of workflows, the level of human oversight, and the trust placed in AI systems. A well-structured human-AI partnership can enhance human agency, but without careful implementation, it can also lead to unintended disempowerment.

This section explores the principles and models of human-AI collaboration, focusing on both the challenges and opportunities they present.

4.1 The Rise of Collaborative Intelligence, Appropriate Trust, and Counteracting Disempowerment

Understanding the varying effectiveness of human-AI collaboration across different tasks is crucial for developing successful collaborative frameworks. Research highlights that AI-human synergy is task-dependent, requiring structured interaction models to ensure optimal performance.

4.1.1 Task-Dependent Nature of Human-AI Collaboration

Recent studies emphasize that human-AI collaboration does not always lead to superior outcomes. Puerta-Beldarrain et al. (2025) found that AI-human synergy is highly task-dependent:

- In **creative tasks**, such as content generation and ideation, AI augments human capabilities, improving efficiency and *offering novel insights*.

- In **decision-making tasks**, particularly those requiring nuanced judgment, AI-human collaboration may reduce performance due to *misalignment in cognitive strategies* and *over-reliance on AI-generated suggestions*.

Vaccaro, Almaatouq, and Malone (2024) further indicate that the effectiveness of AI assistance varies based on user expertise. Novice users often benefit more from AI-guided decision support, while experts may find AI outputs disruptive to their reasoning processes. Felin and Holweg (2024) highlight that AI relies on pattern recognition rather than causal reasoning, meaning that experts may struggle with AI-generated insights that lack explanatory depth. These distinctions highlight the necessity of *task-specific AI integration strategies* that enhance human performance rather than hinder it, ensuring that AI complements rather than disrupts expert workflows.

4.1.2 Essential Factors for Successful Collaboration

Several key factors have emerged as critical determinants for achieving effective human-AI collaboration and maintaining appropriate trust levels.

Key Factors Shaping Human-AI Collaboration:

- **Complementary Strengths**: Humans excel at critical thinking, ethical reasoning, complex problem-solving, and contextual understanding, while AI excels at data processing, pattern recognition, and speed. Felin and Holweg (2024) emphasize that AI lacks causal reasoning, reinforcing the need for *human interpretation* of results.

- **Hybrid Models**: Hybrid Intelligence, emphasizing *iterative interaction and user involvement*, offers a promising model for empowering human experts (Puerta-Beldarrain et al. 2025).

- **Increased Efficiency and Productivity**: AI can automate routine and repetitive tasks, freeing up human experts for *higher-level strategic work*.

- **Improved Accuracy and Decision-Making**: AI provides *data-driven insights* and *reduces human error*, particularly in structured environments.

- **Iterative Interaction**: Ongoing *feedback and communication* are crucial for building trust, refining collaboration, and maintaining human control (Puerta-Beldarrain et al. 2025).

- **Explainability**: AI systems must be *transparent and interpretable*, allowing humans to understand the rationale behind decisions and adjust accordingly (Mehrotra et al. 2024).

4.1.3 Safeguarding Human Agency

As AI systems become more advanced, the risk of human disempowerment grows. Kulveit et al. (2025) warn that improperly designed AI-driven workflows could lead to a *gradual loss of human agency*. To mitigate this risk, AI collaboration must be structured to *enhance human influence rather than replace it*:

- **Maintaining Human Oversight**: Humans must retain meaningful decision-making authority, particularly in *high-stakes applications*. International Telecommunication Union (ITU. 2024) outlines safeguards for ensuring AI decisions remain explainable and contestable.

- **Designing AI for Empowerment**: AI should function as an advisor or co-worker, rather than an autonomous decision-maker, ensuring that humans remain actively engaged (Vaccaro, Almaatouq, and Malone. 2024; Felin and Holweg. 2024).

- **Adaptive AI Interfaces**: AI tools should be designed to *adapt to human reasoning patterns* rather than forcing human users to adjust to machine logic (Lin et al. 2024).

- **Trust Calibration Mechanisms**: Encouraging users to *critically assess AI outputs*, reducing automation bias (Mehrotra et al. 2024; Mitchell et al. 2025; Vaccaro, Almaatouq, and Malone. 2024).

4.1.4 Principles for Balanced Integration

The successful integration of AI systems requires adherence to core principles that prioritize human empowerment while maximizing collaborative benefits:

- **Cognitive Load Consideration**: AI should enhance rather than replace human cognitive effort, as Mehrotra et al. (2024) highlight. If AI assumes too much autonomy, human operators may disengage, increasing the risk of automation complacency (Kulveit et al. 2025).

- **Adaptive Trust Mechanisms**: Over-reliance on AI is a concern in *high-ambiguity environments* (such as medical diagnostics or strategic decision-making). *Structured trust-building mechanisms*, such as confidence scoring, explanations of uncertainty, and interactive AI, help mitigate automation bias (Mehrotra et al. 2024; Mitchell et al. 2025; Vaccaro, Almaatouq, and Malone, 2024; Uuk et al. 2024; Nguyen et al. 2024). It is also important to foster "*appropriate trust*" in AI systems (Mehrotra et al. 2024), ensuring users are neither overly skeptical nor blindly reliant. Data suggests that employees are more likely to trust their own companies to deploy AI responsibly (Mayer et al. 2025), which can be leveraged to build internal trust-building programs.

By incorporating these principles, human-AI collaboration can be structured in a way that *amplifies human expertise rather than diminishing it*. This balanced approach ensures that AI remains a *valuable augmentative tool* rather than a force for displacement.

4.2 Types of Human-AI Integrated Workflows: A Focus on Empowerment

The design of human-AI integrated workflows is crucial for determining whether AI augments human capabilities and preserves agency or contributes to gradual disempowerment (Kulveit et al. 2025). Different workflow types offer varying degrees of human control, involvement, and influence over the decision-making process. Mitchell et al. (2025) provide a valuable framework for understanding these variations by proposing a spectrum of AI agent autonomy, ranging from simple tools with no autonomy to fully autonomous agents capable of independent action and code creation. Understanding where a given workflow falls on this spectrum is crucial for assessing its potential benefits and risks. We categorize these workflows, analyzing their potential for both empowerment and disempowerment.

4.2.1 AI as Assistant

In this model, AI performs routine, time-consuming, or data-intensive tasks, freeing up human experts to focus on higher-level cognitive work, strategic decisions, and tasks requiring uniquely human skills (e.g., critical thinking, creativity, empathy). This type of workflow generally falls on the lower end of the autonomy spectrum, corresponding to Mitchell et al.'s (2025) "Simple Processor" or "Router" levels. The AI is primarily a tool under direct human control.

- **Examples:** AI summarizing legal documents (Christman. 2024), scheduling appointments (Lin et al. 2024), filtering information, automating data entry. Many employees are already using AI as an assistant in their daily work (Mayer et al. 2025), indicating the growing prevalence of this workflow type.

- **Empowerment Potential:** High, if designed correctly. By offloading tedious tasks, AI can *enhance* productivity and allow experts to focus on their core competencies. This can lead to increased job satisfaction and a greater sense of agency.

- **Disempowerment Risk:** If poorly designed, this model can lead to *deskilling* and *over-reliance*. If the AI handles too much of the analytical work, the human expert may become a mere "rubber stamper," losing the opportunity to develop and exercise their own judgment. This relates to the value of *accuracy* (Mitchell et al. 2025), as over-reliance on an assistant with flawed capabilities can lead to errors.

- **Trust Mechanisms:** Transparency and explainability (Mehrotra et al. 2024) are crucial to avoid over-reliance. Users need to understand *why* the AI is performing certain tasks and have the ability to *override* or *adjust* its actions.

4.2.2 AI as Advisor

In this model, while the human expert retains ultimate decision-making authority, the AI provides data-driven insights, recommendations, and predictions to support that process. This increased AI involvement corresponds to a higher level of autonomy, such as the "Tool Call" capability described by Mitchell et al. (2025).

- **Examples:** AI suggesting treatment options in healthcare (Alowais et al. 2023), AI providing risk assessments in finance (Sheharyar. 2024), AI offering legal precedents in law.

- **Empowerment Potential:** Moderate to high. AI can provide valuable information and perspectives that the human expert may not have considered, leading to better-informed decisions.

- **Disempowerment Risk:** Over-reliance on AI recommendations is a significant risk, particularly if the AI's reasoning is opaque or if the human expert lacks the confidence or expertise to critically evaluate the AI's output. This can lead to a gradual erosion of human judgment and expertise. Automation bias can also be a factor (Mehrotra, et. al. 2024). This relates to the values of *truthfulness* and *accuracy* (Mitchell et al. 2025), as reliance on flawed or biased recommendations can lead to poor decisions.

- **Trust Mechanisms:** *Appropriate trust* is paramount (Mehrotra et al. 2024). Experts need to understand the AI's capabilities and limitations, and be able to calibrate their trust accordingly. *Explainability* is key: understanding *why* the AI is making a particular recommendation is crucial for informed decision-making. Mechanisms for *challenging* or *overriding* AI recommendations are also essential for maintaining human agency.

4.2.3 AI as Co-worker

This model emphasizes *iterative interaction* and *shared control* between humans and AI. Humans and AI work together on tasks, each contributing their unique skills and expertise, and actively participating in the learning loop. This is best exemplified by the *Hybrid Intelligence* approach (Puerta-Beldarrain et al. 2025). This workflow type typically involves higher levels of AI autonomy, such as Mitchell et al.'s (2025) "Multi-Step Agent" level, where the AI can plan and execute sequences of actions. Lin et al. (2024) provide concrete examples of this in their "decision-oriented dialogue" tasks, such as collaborative itinerary planning or multi-party meeting scheduling. Also, Nguyen et al. (2024) analyzed how students employed strategies in academic writing using generative AI. All these studies show how AI systems and humans can collaborate and co-work in iterative tasks.

- **Examples:** The air quality awareness application (Guo et al. 2022), where users actively provide feedback, customize the system, and participate in the data collection and

analysis process. Co-writing with AI, where both contribute to the creative process (Boussioux et al. 2024).

- **Empowerment Potential:** High. This model offers the greatest potential for *genuine collaboration* and *mutual learning*. By actively participating in the AI's learning process, human experts can shape the system's behavior and ensure it aligns with their goals and values.

- **Disempowerment Risk:** Lower than other models, if designed correctly. The key is to ensure that the human retains meaningful control and agency within the collaborative process. Poorly designed interfaces or overly complex AI systems could still lead to disempowerment. Mitchell et al. (2025) emphasize that even at this level, risks related to *safety*, *security*, and *truthfulness* remain, and can be amplified by the increased autonomy.

- **Trust Mechanisms:** Trust is built through iterative interaction, feedback, and shared control. Transparency and explainability are important, but active participation is the primary driver of trust (Mehrotra et al. 2024; Puerta-Beldarrain et al. 2025).

4.2.4 AI as Manager

In this model, the AI system has significant autonomy in decision-making, with humans playing a limited oversight role. Humans may be involved in setting high-level goals or constraints, but the AI makes the majority of operational decisions. Alternatively, the human may be the executor of the low-level tasks. This corresponds to the highest levels of autonomy on the Mitchell et al. (2025) scale, approaching or achieving "Fully Autonomous Agent" status. This is the level where Mitchell et al. (2025) express the strongest concerns about potential risks.

- **Examples:** Fully autonomous vehicles, AI-driven trading systems, AI managing complex infrastructure.

- **Empowerment Potential:** Low, unless very carefully designed. This model carries the highest risk of *gradual disempowerment*.

- **Disempowerment Risk:** High. If human oversight is insufficient or ineffective, AI systems could make decisions that are harmful, unethical, or contrary to human values. The "loss of control" systemic risk (Uuk et al. 2024) is most prominent in this model. Even with oversight, there may be cases where humans cannot intervene fast enough.

- **Trust Mechanisms:** Extremely high levels of *reliability, transparency, and accountability* are required. Robust mechanisms for *human intervention and override* are essential. *Ethical safeguards* and *value alignment* are paramount. This model requires the *most rigorous* governance and oversight.

The design of human-AI integrated workflows is crucial in determining whether AI serves to augment human capabilities and preserve agency or gradually contributes to *disempowerment* (Kulveit et al. 2025). Different workflow models offer varying degrees of human control, involvement, and decision-making influence.

While lower-autonomy AI systems (such as Assistants and Advisors) can enhance productivity and provide valuable insights, higher-autonomy systems (such as AI as Co-worker or AI as Manager) introduce significant risks that require careful trust mechanisms and human oversight (Mitchell et al. 2025). The gradual nature of disempowerment means that seemingly benign AI

applications could, over time, contribute to a long-term erosion of human decision-making authority (Kulveit et al. 2025).

To address these challenges, it is critical to evaluate AI systems across the full spectrum of autonomy, as proposed by Mitchell et al. (2025). Moving forward, AI workflows must be designed to preserve meaningful human control and agency, ensuring that AI functions as a tool for augmentation rather than displacement (Mitchell et al. 2025).

5. The Transformation of Expert Roles: Embracing Change and Maintaining Agency

The rise of AI is not simply automating tasks; it is fundamentally reshaping the roles and responsibilities of human experts. This transformation requires a shift in mindset, a focus on developing new skills, and a proactive approach to maintaining agency in an AI-driven world. The traditional role of the expert as a *repository of knowledge* is becoming less valuable as AI systems gain the ability to access, process, and analyze vast amounts of information (Boussioux et al. 2024). Instead, the future of expertise lies in *leveraging uniquely human capabilities* and strategically integrating AI to *augment* those capabilities. Given that many companies are still early in their AI adoption journey (Mayer et al. 2025), experts are needed to guide organizations towards maturity and value creation.

5.1 From Knowledge Repository to Knowledge Synthesizer, Strategist, Overseer, and Ethical Arbiter

AI's capacity to access, process, and analyze vast amounts of information far exceeds human capabilities. This is shifting the value proposition of human expertise. Simply *possessing* knowledge is becoming less important than the ability to *synthesize* information from multiple sources (including AI), apply *critical thinking* and *contextual understanding* (Boussioux et al. 2024; Felin and Holweg 2024), make *nuanced judgments*, and provide *strategic direction*.

Key themes

- **Shift in Value:** The value of human expertise is shifting from simply possessing knowledge to the ability to synthesize information, apply critical thinking, and make nuanced judgments. Experts will become *interpreters* of AI-generated insights, *contextualizing* them within broader organizational, ethical, and societal frameworks, and acting as *strategists* to guide action based on this comprehensive understanding.

- **Lifelong Learning:** Continuous learning and skill development are no longer optional; they are *essential* for remaining relevant and competitive. Experts must embrace a "growth mindset" and actively seek opportunities to upskill and reskill.

- **Focus on Higher-Order Skills:** Critical thinking, complex problem-solving, creativity, communication, emotional intelligence, and ethical reasoning will be increasingly valuable. These are the skills that are most difficult for AI to replicate.

- **New Roles:** Experts will increasingly take on roles such as:

- **AI Managers:** Overseeing the development, deployment, and operation of AI systems.

- **AI Interpreters:** Translating AI-generated insights into actionable strategies and recommendations.

- **AI Validators:** Critically evaluating AI outputs, identifying biases, and ensuring accuracy.

- **AI Trainers:** Providing feedback to AI systems and participating in their learning process (Puerta-Beldarrain et al. 2025). This iterative process is also crucial for building trust (Puerta-Beldarrain et al. 2025; Mehrotra et al. 2024).

- **Ethical Arbiters:** Making ethical judgments that AI systems cannot, particularly in complex or ambiguous situations where human values and societal norms are paramount. This includes addressing the systemic risks outlined by Uuk et al. (2024).

5.2 The Importance of Human Oversight, Ethical Considerations, Control, and Preventing Disempowerment

Human oversight remains absolutely essential, not just for accuracy and accountability, but for *actively preventing gradual disempowerment* (Uuk et al. 2024; Kulveit et al. 2025). As AI systems become more capable and autonomous, the risk of eroding human influence and control increases. This risk is particularly pronounced as AI systems move along the autonomy spectrum towards greater independence (Mitchell et al. 2025).

Key themes

- **Accountability and Responsibility:** Humans must retain ultimate responsibility for decisions made and actions taken, even when those decisions are informed or assisted by AI. Clear lines of authority and accountability must be established.

- **Bias Detection and Mitigation:** Human oversight is crucial for identifying and addressing biases in AI algorithms and datasets. This requires a deep understanding of both the technical aspects of AI and the social and ethical implications of its use.

- **Ethical Decision-Making:** Human judgment is needed to address the complex ethical dilemmas that arise from the use of AI, particularly in high-stakes domains. This includes considering the potential for AI to be used for malicious purposes (e.g., misinformation, manipulation).

- **Explainability and Transparency:** Understanding *how* AI systems make decisions is essential for trust, accountability, and effective oversight. Experts must be able to explain AI-generated insights to stakeholders and justify decisions based on those insights (Mehrotra et al. 2024).

- **Maintaining Control:** Mechanisms are needed to prevent "loss of control" over AI systems (Uuk et al. 2024), and to ensure that humans retain the ability to intervene and *override* AI decisions when necessary. This is crucial for preventing runaway automation and ensuring that AI remains aligned with human values. This aligns with the core

argument of Mitchell et al. (2025), who advocate against the development of fully autonomous AI agents precisely because of the risk of irreversible loss of control.

- **Preventing Disempowerment:** Oversight must be structured to *actively prevent* the erosion of human influence and control. This requires:

    - **Clear Lines of Authority and Responsibility:** Ensuring that humans have the final say in critical decisions.

    - **Mechanisms for Human Intervention and Override:** Providing ways for humans to intervene and override AI decisions when necessary.

    - **Ongoing Monitoring of AI System Behavior and Impact:** Continuously assessing the impact of AI on human agency and control.

    - **Proactive Identification and Mitigation of Systemic Risks:** Addressing the risks outlined by Uuk et al. (2024), such as discrimination, power concentration, and irreversible change.

    - **Promoting Hybrid Intelligence:** Ensuring the appropriate balance of power.

6. Human Resilience Strategies in the AGI Era: Preparing for an Uncertain Future

The rapid advancement of AI, particularly the development of increasingly autonomous systems (Mitchell et al. 2025), presents both unprecedented opportunities and significant challenges for specialized professions. While AI can automate tasks, enhance decision-making, and improve efficiency, it also poses systemic risks (Uuk et al. 2024) and the threat of *gradual disempowerment* (Kulveit et al. 2025), where human influence and control over critical systems could be eroded over time. This is a topic of growing debate (Pew Research Center, 2023)**.** To navigate this uncertain future, individuals, organizations, and societies must prioritize human resilience strategies that focus on lifelong learning, effective human-AI collaboration, and proactive policy interventions. The increasing investment in AI (Mayer et al. 2025) underscores the urgency of developing these resilience strategies.

6.1 Lifelong Learning and Skill Development: The Adaptive Expert

Continuous learning and skill development are no longer optional; they are *essential* for professionals to remain relevant and adaptable in an AI-driven world. The traditional role of the expert as a repository of knowledge is becoming less valuable as AI systems gain the ability to access, process, and analyze vast amounts of information (Boussioux et al. 2024). The future of expertise lies in leveraging *uniquely human capabilities* and strategically integrating AI to *augment* those capabilities.

Key skills for the future workforce include:

- **Critical Thinking and Complex Problem-Solving:** The ability to analyze complex situations, evaluate AI-generated insights (Felin and Holweg, 2024), and make informed decisions is paramount. This is particularly important in mitigating the risks associated

with higher levels of AI autonomy, where human oversight and judgment remain crucial (Mitchell et al. 2025).

- **Creativity and Innovation:** As AI excels at tasks requiring pattern recognition and optimization, uniquely human skills like creativity and innovation become increasingly valuable (Boussioux et al. 2024).

- **Emotional Intelligence and Communication:** Effective collaboration, leadership, and ethical decision-making require strong emotional intelligence and communication skills.

- **AI Literacy:** A fundamental understanding of AI capabilities, limitations, and ethical implications is essential for all professionals, not just AI specialists. This includes understanding the potential for bias, hallucinations, and other failure modes in AI systems (Lin et al. 2024; Mitchell et al. 2025), as well as the importance of appropriate trust (Mehrotra et al. 2024).

- **Data Literacy:** The ability to collect, clean, prepare, interpret, and use data is crucial in an increasingly data-driven world.

- **Adaptability and Flexibility:** The ability to learn new skills and adapt to changing job requirements will be crucial, as AI reshapes the nature of work.

To foster these skills, governments, organizations, and individuals must invest in:

- **Education and Training:** Develop curricula and training programs that prioritize critical thinking, creativity, problem-solving, and AI literacy.

- **"Meta-Skills":** Focus on developing "meta-skills" like learning how to learn, adaptability, and resilience.

- **A Culture of Lifelong Learning:** Encourage and support continuous learning and skill development throughout individuals' careers.

6.2 Embracing Human-AI Collaboration: Becoming the "AI-Augmented Expert"

The future of work is not about humans *versus* AI, but about humans *collaborating with* AI. This requires developing the skills, mindset, and strategies to effectively leverage AI's capabilities while maintaining human agency and control. Key skills for effective human-AI collaboration include:

- **Understanding AI Capabilities and Limitations:** Knowing when to trust AI and when to rely on human judgment is essential. This requires a deep understanding of both the strengths and weaknesses of AI systems.

- **Developing Appropriate Trust:** Learning to calibrate trust in AI systems based on their performance and limitations (Mehrotra et al. 2024), and understanding how to utilize *explainability* features (when available) to assess the reasoning behind AI outputs.

- **Effective Communication and Interaction:** Learning how to communicate with AI systems, provide feedback, and interpret their outputs. This includes developing strong *prompting skills* to effectively guide AI behavior (Lin et al. 2024).

- **Adapting to New Workflows:** Being flexible and adaptable to new ways of working that integrate human and AI contributions. This may involve re-designing workflows to leverage the complementary strengths of humans and AI.

- **Active Participation in the AI's Learning Loop:** Providing feedback, correcting errors, and guiding the AI's learning process (Puerta-Beldarrain et al. 2025). This active participation is a key element of fostering appropriate trust and preventing over-reliance (Mehrotra et al. 2024). It also helps to ensure that AI systems align with human values and goals. Different interaction patterns and collaboration styles may emerge, such as the "centaur" and "cyborg" approaches (Dell'Acqua et al. 2024), highlighting the need for research into optimal human-AI teaming strategies.

To promote effective human-AI collaboration, organizations should:

- **Invest in Training:** Develop training programs that focus specifically on human-AI collaboration skills, including prompting, feedback provision, and workflow adaptation.

- **Foster Knowledge Sharing:** Encourage experienced professionals to mentor others in adapting to AI and integrating it into their work.

- **Promote a "Growth Mindset":** Foster a culture that embraces learning, experimentation, and adaptation.

6.3 Economic and Policy Strategies for the Future

Addressing the potential economic and social impacts of AI-driven changes in the labor market requires a multi-faceted approach, encompassing education, social safety nets, economic policies, and regulatory frameworks. These strategies must be proactive and adaptable to the evolving landscape of human-AI interaction. The "*AI governance paradox*" (ITU. 2024), where technological capabilities outpace governance and oversight tools development, needs to be taken into account:

Key strategies include:

- **Rethinking Education Systems:** Future generations must be prepared for an AI-driven world. This requires a shift in educational priorities, emphasizing skills that complement AI, such as critical thinking, creativity, complex problem-solving, and emotional intelligence. Furthermore, AI and data literacy should be prioritized for all citizens (Pew Research Center. 2023), not just for specialists.

- **Social Safety Nets:** Policymakers should consider implementing social safety nets to address potential job displacement caused by AI automation. Universal basic income (UBI) and other innovative social support mechanisms may become increasingly necessary to ensure a basic standard of living for those whose jobs are displaced by AI.

- **Wealth Redistribution Mechanisms:** The potential for AI to exacerbate existing inequalities necessitates a discussion about wealth redistribution mechanisms. Policies such as wealth taxes, progressive taxation, and employee ownership models could help

to mitigate the concentration of wealth and power in the hands of those who control AI technologies.

- **Promoting Human-Centric Industries:** Investing in sectors that are less susceptible to automation and that leverage uniquely human capabilities, such as caregiving, education, and the arts, can help to create new job opportunities and maintain a balance between human and AI contributions.

- **Regulations on Decision-Making Authority:** As AI systems move towards higher levels of autonomy on the spectrum described by Mitchell et al. (2025), it becomes increasingly important to implement policies that limit the delegation of critical decisions to AI and mandate human review in high-stakes situations (Kulveit et al. 2025). This is particularly crucial in domains where errors could have severe consequences, such as healthcare, finance, and criminal justice. To maintain human oversight, robust mechanisms for control should include Human-in-the-Loop (HITL) and Human-on-the-Loop (HOTL) frameworks.

  - **HITL** ensures human intervention before AI finalizes decisions, often applied in cybersecurity, high-risk automation, and certain aspects of fraud detection (Mosqueira-Rey et al. 2023; Maathuis 2024; Uuk et al. 2024).

  - **HOTL** allows AI to operate with greater autonomy while ensuring human oversight and override capabilities, commonly used in AI-assisted diagnostics, strategic defense AI operations, and financial market governance (Mehrotra et al. 2024; ITU. 2024; Credo.ai. n.d.).

- **Addressing Systemic Risks:** Proactive measures are needed to identify and mitigate the systemic risks associated with AI, as outlined by Uuk et al. (2024). This includes addressing issues such as loss of control, discrimination, economic disruption, and irreversible societal changes.

- **International Cooperation:** Given the global nature of AI development and deployment, international cooperation is essential (ITU. 2024). This includes developing international standards for AI governance, focusing on interoperability, shared values, and ethical principles, and promoting collaboration and knowledge sharing among nations, particularly to support developing countries (ITU. 2024). Addressing the "AI divide" and ensuring equitable access to the benefits of AI is a crucial aspect of global governance.

- **Transparency and Explainability**: Addressing the "AI governance paradox" (ITU. 2024) requires investment in research and development of tools for monitoring, checking, and controlling AI systems, as well as ensuring transparency and explainability in AI systems, particularly in high-stakes domains (Mehrotra et al. 2024; Pew Research Center. 2023).

These strategies emphasize a balanced approach, ensuring AI fosters economic growth while safeguarding human agency, ethical standards, and societal well-being.

6.4 Fostering Uniquely Human Strengths

While AI systems excel at data processing, pattern recognition, and optimization, certain human capabilities remain difficult, if not impossible, to replicate. Fostering these uniquely human

strengths is essential for maintaining a balanced and thriving society in an AI-driven world. These strengths include:

Key strategies include:

- **Creativity and Innovation:** While AI can assist in creative tasks (Ivcevic and Grandinetti 2024), true innovation often requires the ability to think outside existing paradigms, make unexpected connections, and generate genuinely novel ideas. However, it is important to investigate whether AI models can achieve this, as some research shows that they can produce novel ideas and expand creative problem-solving through parallel exploration (Boussioux et al. 2024). Additionally, human involvement remains essential in guiding AI's creative outputs, refining ideas, and integrating insights into meaningful applications (Dell'Acqua et al. 2024).

- **Adaptability to new situations:** While AI can be very usefull, it is crucial for humans to know how to deal with situations where there is no AI, or AI is failing.

- **Ethical Reasoning and Judgment:** Humans possess a capacity for moral reasoning and ethical judgment that goes beyond simply following rules or maximizing predefined objectives. This is particularly important in complex and ambiguous situations where values and societal norms must be considered.

- **Emotional Intelligence and Empathy:** Understanding and responding to the emotions of others, building relationships, and providing care and support are crucial aspects of human interaction that are difficult for AI to replicate.

- **Contextual Understanding and Common Sense:** Humans have a deep understanding of the world and can apply common sense reasoning to navigate complex and unpredictable situations. This ability to contextualize information and make nuanced judgments is often lacking in AI systems.

Investing in education, training, and cultural practices that cultivate these uniquely human strengths is crucial for ensuring that humans remain active and valuable contributors to society, even as AI takes on an increasingly prominent role.

7. Case Studies of Successful AI-Human Collaboration

This section presents case studies that illustrate successful human-AI collaboration, with a particular focus on how these collaborations *empower* human experts and maintain human agency. We examine the AI application, the collaborative workflow, and the resulting benefits, challenges, and trust mechanisms. We also analyze how these examples align with the principles of human-centered AI and the mitigation of potential risks, such as gradual disempowerment (Kulveit et al. 2025). These examples also highlight the importance of moving beyond localized pilot projects towards broader implementation (Mayer et al. 2025).

7.1 Hybrid Intelligence in Air Quality Awareness

- **Contextual Understanding and Common Sense:** Humans have a deep understanding of the world and can apply common sense reasoning to navigate complex and unpredictable situations. This ability to contextualize information and make nuanced judgments is often lacking in AI systems.

- **AI Application:** A system designed to raise awareness about air quality, likely using sensor data and predictive modeling.

- **Collaborative Workflow:** As described by Puerta-Beldarrain et al. (2025), this case embodies a *Hybrid Intelligence* approach. Unlike traditional top-down models in which people passively receive AI outputs, users actively shape the system's learning and decision-making. Specifically:

  - **Iterative Interaction:** Users provide ongoing feedback about predictions, contribute data (e.g., local sensor readings), and adapt how the system presents information.

  - **Shared Control:** AI handles sensor-data processing and pattern recognition, while humans bring contextual knowledge, ethical judgment, and real-world experience.

- **Benefits:** This Hybrid Intelligence system leads to higher user engagement and motivation, improved data quality through user feedback (as individuals can spot anomalies or nuanced issues in local measurements), and a stronger sense of system ownership that encourages sustainable use and fosters trust.

- **Challenges:** Key challenges involve designing interfaces that enable meaningful participation from non-technical users, ensuring that the AI model adjusts to evolving user goals and constraints, and striking a workable balance between automation and direct human oversight.

- **Trust Mechanisms:** Trust is cultivated through frequent, iterative interactions that allow humans to observe and shape the system's outputs, transparent explanation of how air-quality predictions are generated, and decision-making processes in which users retain genuine influence.

- **Empowerment Analysis:** This application illustrates how Hybrid Intelligence can actively engage community members. Rather than simply broadcasting sensor outputs, the system invites users to co-develop insights, thus preserving human agency. The approach also aligns with human-centered AI practices by keeping humans at the core of system evolution, limiting the risk of disempowerment.

7.2 Decision-Oriented Dialogue

- **AI Application:** AI large language models (LLMs) (e.g., GPT-4) serve as dialogue partners in three tasks.

- **Collaborative Workflow:** Lin et al. (2024) propose a *decision-oriented dialogue* framework, highlighting how humans can interact with LLMs to co-construct solutions. Three example tasks are central to their study:

  - **Assignment:** The AI assists in matching papers to suitable reviewers. Through conversation, the human clarifies constraints or priorities (e.g., expertise, availability), while the AI synthesizes available data to suggest assignments.

  - **Planning:** As a travel assistant, the AI helps the human plan an itinerary by providing potential routes, costs, and schedules. The user refines these ideas by adding preferences and constraints.

- - **Mediation:** The AI coordinates group scheduling by interacting with multiple human participants separately, then proposing a compromise schedule.

  In all of the tasks, the AI system and the humans must interact and collaborate with the aim of accomplishing a higher and shared goal. The human can provide initial data and prompts, as well as refine the generated ideas.

- **Benefits:** The primary benefits include rapid access to extensive datasets, such as reviewer databases or flight and hotel listings, and fast generation of different plan variations or scheduling options.

- **Challenges:** These dialogues face obstacles that include the AI's propensity to produce "hallucinations" (i.e., unsupported facts), as well as limitations in understanding real-world nuances, user motivations, or interpersonal group dynamics.

- **Trust Mechanisms:** Trust is promoted through goal-driven exchanges in which both humans and the AI articulate their reasoning, along with explicit human confirmations or rejections (e.g., "Yes, let's select these reviewers" or "No, that's not feasible").

- **Empowerment Analysis:** Human participants retain final decision-making authority. The AI can handle data-heavy tasks, but humans supply strategic judgment, preferences, and ethical considerations. This preserves autonomy and supports collaborative performance gains.

7.3 AI-Assisted Writing

- **AI Application:** Generative AI tools (e.g., ChatGPT, Google Bard) assist graduate students with academic writing tasks.

- **Collaborative Workflow:** Nguyen et al. (2024) examine how students integrate AI tools into their writing process. Rather than being locked into a single workflow, students can:

  - Generate an initial draft or outline.

  - Brainstorm or refine ideas.

  - Edit and reorganize existing text.

  - Check grammar and style.

  - Explore references or find alternative perspectives.

  A key point is identifying how *self-regulated learning (SRL)* strategies emerge when humans collaborate with AI.

- **Benefits:** Key benefits include enhanced writing efficiency and the possibility of better-structured drafts, overcoming writer's block through quick suggestions of new angles, and exposure to a variety of phrasing or argumentation styles that can broaden learners' writing skills.

- **Challenges:** Potential issues arise from excessive reliance on AI outputs without careful vetting for correctness or coherence, as well as the erosion of independent writing skills if learners simply accept AI-generated text.

- **Trust Mechanisms:** Trust is reinforced by transparent guidelines that clarify appropriate AI usage (e.g., how to cite AI-assisted content) and by encouraging students to maintain a critical mindset when verifying sources, checking factual accuracy, and refining the AI's suggestions.

- **Empowerment Analysis:** Students who treat the AI as a supportive tool—actively critiquing and building upon generated drafts—can enhance both their writing and metacognitive skills. By contrast, passively accepting AI content risks diminishing a learner's ownership and skill development. The researchers highlight that explicit self-regulated learning strategies help students maintain an active role and develop a more productive human–AI collaboration.

8. Policy Recommendations for Ensuring a Balanced AI-Driven Future

The rapid advancement and deployment of AI technologies necessitate a proactive and comprehensive approach to policy-making. The goal is to harness the benefits of AI while mitigating the risks, particularly the potential for *gradual disempowerment* (Kulveit et al. 2025) and the erosion of human agency. Policy recommendations must address multiple levels, from individual skills and organizational practices to national regulations and international cooperation. A balanced approach will recognize the diverse perspectives and needs of stakeholders, promoting both innovation and responsible development (Pew Research Center. 2023), and must take into account the "*AI governance paradox*" (ITU. 2024).

8.1 Investing in Education and Training for the AI Era

A foundational element of a balanced AI-driven future is a workforce equipped to navigate the changing landscape. This requires a fundamental shift in education and training, prioritizing AI and data literacy for all citizens (Pew Research Center, 2023). Curricula should not only cover the technical aspects of AI but also its capabilities, limitations, and ethical implications (Mehrotra et al. 2024). Specifically, individuals need to be aware of potential biases, hallucinations, and other failure modes that can arise in AI systems (Lin et al. 2024; Mitchell et al. 2025). Beyond foundational literacy, specialized training programs are needed to equip professionals with the skills to *effectively collaborate* with and *oversee* AI systems. This includes developing expertise in areas like prompt engineering (Lin et al. 2024), providing feedback, adapting to new workflows, and actively participating in the AI's learning loop (Puerta-Beldarrain et al. 2025).

Furthermore, a culture of lifelong learning must be fostered, encouraging continuous upskilling and reskilling to keep pace with the rapid evolution of AI technologies. This includes an emphasis on meta-skills such as learning how to learn, adaptability, and resilience (Puerta-Beldarrain et al. 2025).

8.2 Promoting Responsible AI Development and Deployment

The rapid advancement of AI demands a governance framework that is adaptive, transparent, and aligned with human interests. AI's systemic risks (Uuk et al. 2024) must be directly addressed through tailored regulatory measures to ensure AI serves as a force for augmentation rather than disempowerment.

8.2.1 Task-Dependent Nature of Human-AI Collaboration

Effective AI governance requires a multi-pronged approach, combining ethical guidelines, regulatory frameworks, and oversight mechanisms. Ethical guidelines must emphasize human well-being, fairness, and accountability, particularly in *high-stakes domains* where the potential for harm is greatest (Mehrotra et al. 2024).

A critical principle is maintaining *human oversight* and *control* through:

- **HITL and HOTL frameworks**: As defined in Section 6.3, these approaches ensure that AI systems incorporate human supervision at different stages of operation, reinforcing accountability and trust (Maathuis 2024; Uuk et al. 2024; Mehrotra et al. 2024; ITU 2024).

- **Explainability requirements**: AI systems must be transparent, allowing human operators to audit and challenge AI-generated decisions to prevent reliance on opaque decision-making models (Mehrotra et al. 2024).

- **Bias detection and mitigation**: Regulatory mandates requiring AI audits for fairness, particularly in finance, healthcare, and hiring, to prevent discriminatory outcomes and algorithmic biases (Uuk et al. 2024).

8.2.2 Mapping Systemic Risks to Policy Solutions

Each systemic risk outlined earlier aligns with specific regulatory interventions:

- **Loss of Control → Human-in-the-loop (HITL) requirements**: To address the risks of loss of control, regulatory measures should ensure meaningful human oversight in high-stakes environments such as law enforcement and critical infrastructure. This aligns with concerns about the gradual erosion of human agency, as AI systems become increasingly embedded in decision-making processes (Kulveit et al. 2025).

- **Democracy Risks → Transparency and accountability mandates**: To mitigate democracy risks, regulatory frameworks must enforce transparency, accountability, and oversight in AI-generated content, decision automation, and algorithmic governance. Research highlights the need for governance structures to prevent AI-driven erosion of democratic norms (Uuk et al. 2024).

- **Economic Disruption → Workforce adaptation policies**: To address economic disruption, governments must implement reskilling initiatives and workforce adaptation strategies that equip professionals to work alongside AI systems. Studies highlight the need for ongoing training to ensure effective human-AI collaboration rather than workforce displacement (Mehrotra et al. 2024).

- **Irreversible Change Risks → AI Moratoriums on unsafe deployments**: To prevent irreversible societal risks, regulatory frameworks should impose strict governance on fully autonomous AI deployments, particularly in critical areas such as military

command, legal rulings, and healthcare. Research suggests that human oversight remains indispensable to ensuring safety and accountability in AI-driven decisions (Mitchell et al. 2025).

By aligning governance strategies with identified risks, AI policies can balance innovation, accountability, and human empowerment while mitigating existential threats. Ensuring AI remains an augmentative force rather than a destabilizing one requires *proactive oversight* and *adaptable policy frameworks* that leverage HITL and HOTL models effectively.

8.3 Supporting Workforce Adaptation

The potential economic and social impacts of AI-driven changes in the labor market require proactive strategies to support workforce adaptation. This includes providing resources and support for professionals to upskill and reskill, enabling them to transition to new roles and industries. Policymakers should also consider social safety nets, such as universal basic income (UBI) or other support mechanisms, to address potential job displacement. Furthermore, exploring wealth redistribution mechanisms, such as wealth taxes, progressive taxation, or employee ownership models, may be necessary to mitigate increasing inequality. Investing in human-centric industries that leverage uniquely human capabilities, such as caregiving, education, and the arts, can create new job opportunities and promote a more balanced economy.

8.4 Cultivating a Culture of Human-AI Collaboration and Empowerment

Ultimately, a balanced AI-driven future requires a shift in mindset, from viewing AI as a replacement for human capabilities to recognizing it as a *tool for augmentation and collaboration*. Promoting a positive and collaborative attitude towards AI, emphasizing its potential to enhance human capabilities, is crucial.

This involves encouraging the development of *Hybrid Intelligence* workflows, which leverage the complementary strengths of humans and AI, allowing for iterative interaction, shared control, and active user participation (Puerta-Beldarrain et al. 2025; Lin et al. 2024; Dell'Acqua et al. 2024).

Fostering a culture of continuous learning, adaptation, and ethical reflection is essential for navigating the evolving landscape of human-AI interaction (Puerta-Beldarrain et al. 2025).

8.5 Enhancing Uniquely Human Capabilities

8.5.1 The Critical Need for Active Human Engagement

The maintenance of human capabilities in an AI-driven world requires more than just robust governance and education; it demands active exercise of our uniquely human abilities.

Even with robust AI governance, investment in education, and human-centered design principles, there remains a critical need to ensure that humans continue to actively exercise their uniquely human capabilities. The risk is not simply that AI will make bad decisions, but that humans will become overly reliant on AI, leading to a degradation of skills and a diminished capacity for independent thought and action. We need to focus on the importance of maintaining human

capabilities through active engagement and challenge, going beyond simply having the skills in the first place. It's a "use it or lose it" principle applied to human intellect and agency.

8.5.2 The Risk of Skill Degradation

The increasing reliance on AI systems poses risks to human cognitive engagement and independent decision-making. While AI can undoubtedly augment human capabilities and improve efficiency, there is a danger that over-reliance on AI recommendations may reduce users' ability to critically assess information, leading to potential loss of judgment confidence and a decline in independent problem-solving behaviors.

Studies have identified automation bias as a key factor contributing to these risks. If humans routinely defer to AI-generated decisions without critical evaluation, they may develop a habit of unquestioning acceptance, leading to inappropriate reliance on AI even in cases where human judgment is preferable. Research has shown that misplaced trust in automation can affect users' ability to exercise oversight and correction, reinforcing over-reliance on AI systems (Mehrotra et al. 2024).

8.5.3 Key Strategies for Maintaining Human Agency

Preserving human capabilities requires deliberate strategies that go beyond traditional education and training approaches.

Therefore, strategies for maintaining human agency must go beyond simply providing opportunities for education and training. They must also create environments and incentives that encourage humans to:

- **Actively engage with complex problems:** This means resisting the temptation to always take the easiest path offered by AI. It requires seeking out challenges that require human ingenuity, creativity, and critical thinking.

- **Exercise independent judgment:** Even when AI provides recommendations or insights, humans should be encouraged to critically evaluate those outputs, question the underlying assumptions, and make their own informed decisions.

- **Embrace "productive struggle":** Learning and growth often require effort and even discomfort. AI systems should be designed to support this process, not to eliminate it entirely.

- **Seek out diverse perspectives:** AI systems, even those trained on vast datasets, can reflect the biases and limitations of their training data. Humans must actively seek out diverse perspectives and challenge the assumptions embedded in AI systems.

8.5.4 Practical Implementation Approaches

Several concrete methods can help maintain and enhance human capabilities in an AI-augmented environment.

Examples of how this might be achieved in practice include:

- **Iterative Adversarial Testing:** Actively testing AI systems through adversarial approaches, probing their weaknesses, and identifying potential failure modes aligns with iterative AI testing methods discussed in dialogue research (Lin et al. 2024).

- **Gamification of Complex Tasks:** Designing engaging and challenging tasks that require human ingenuity and creativity, even if AI could technically perform those tasks more efficiently.

- **Encouraging Reflective Decision-Making:** Encouraging deliberate, reflective decision-making can help maintain critical thinking in AI-augmented workflows. While efficiency is often prioritized, structured reflection is essential for mitigating automation bias and ensuring ethical decision-making.

- **Hybrid Approaches**: Creating systems where humans define relevance while AI assists in analysis and suggesting strategies (Felin and Holweg. 2024).

8.5.5 Balancing Individual Choice and Societal Needs

While respecting individual autonomy in AI adoption, society must ensure the preservation of human cognitive capabilities.

It is important to acknowledge that some individuals may *choose* to offload tasks to AI, even if it means losing some skills. This is a personal choice, and there may be valid reasons for doing so. However, at a societal level, it is crucial to ensure that there are ample opportunities for humans to *exercise* their cognitive abilities and to maintain a high level of engagement with the world around them.

In essence, maintaining human agency in an AI-driven world requires a conscious and deliberate effort to cultivate and preserve uniquely human capabilities. This is not just about preventing the negative consequences of over-reliance on AI; it's about ensuring that humans continue to thrive, innovate, and contribute to a meaningful future.

8.6 Qatar-Specific Policy Considerations

Qatar's commitment to harnessing Artificial Intelligence (AI) is evident through its strategic initiatives and policies. To further align with global AI governance frameworks and address local priorities, the following policy recommendations are proposed:

8.6.1 Data Governance and Sharing

- **Establish a Data Strategy Office**: Create a dedicated body responsible for educating organizations on the strategic use of data and developing guidelines that encourage data sharing while safeguarding privacy. This initiative aims to democratize AI development and foster innovation (QCAI. 2019).

- **Promote Multilateral Data-Sharing Agreements**: Lead diplomatic efforts to establish international data-sharing agreements, especially among countries with small populations, to enhance AI application development and overcome data scarcity challenges (QCAI. 2019).

8.6.2 Ethical and Transparent AI Deployment

- **Implement Comprehensive AI Ethics Guidelines**: Ensure that AI systems adhere to ethical standards emphasizing human oversight, transparency, and accountability. Regular audits and assessments should be conducted to maintain public trust and compliance with ethical norms (MCIT. 2024).

- **Enhance Data Privacy Measures**: Strengthen data protection laws to ensure that AI applications comply with privacy standards, thereby safeguarding individual rights and fostering public trust in AI technologies (MCIT. 2024).

8.6.3 Capacity Building and Talent Development

- **Integrate AI into Educational Curricula**: Embed AI literacy and skills development into all levels of education, from K-12 to higher education, to prepare a workforce capable of thriving in an AI-driven economy (QCAI. 2019).

- **Attract and Retain AI Talent**: Develop policies offering competitive incentives, streamlined residency processes, and robust research opportunities to draw top AI professionals globally, enhancing Qatar's position as an AI innovation hub (QCAI. 2019).

8.6.4 Infrastructure and Research Development

- **Invest in AI Research and Development**: Allocate resources to AI R&D, focusing on areas such as natural language processing, computer vision, and AI ethics, to drive innovation and address sector-specific challenges (QCAI. 2019).

- **Enhance Digital Infrastructure**: Continue to develop robust digital infrastructure, including data centers and high-speed connectivity, to support AI applications and attract multinational technology companies (International Trade Administration. 2024).

8.6.5 Regulatory Frameworks and International Collaboration

- **Establish AI Regulatory Sandboxes**: Create controlled environments that allow for the testing and development of AI applications under regulatory supervision, balancing innovation with ethical and legal compliance (NCSA. 2024).

- **Engage in International AI Partnerships**: Collaborate with global AI entities to share knowledge, establish best practices, and harmonize standards, ensuring that Qatar remains at the forefront of AI advancements (The Peninsula. 2025).

By implementing these recommendations, Qatar can strengthen its AI ecosystem, ensuring that AI technologies are developed and deployed in a manner that aligns with national values and contributes to sustainable economic growth.

9. Conclusion

The rapid advancement of AI presents both unprecedented opportunities and significant challenges for specialized professions. While AI can automate tasks, enhance decision-making, and improve efficiency, it also poses *systemic risks* (Uuk et al. 2024) and the threat of *gradual disempowerment* (Kulveit et al. 2025), where human influence and control over critical systems could be eroded over time. The future of expertise is not about replacement by AI, but about *strategic integration and augmentation*.

Key Takeaways:

- **Proactive Adaptation**: Professionals must embrace lifelong learning, develop skills that complement AI, and learn to collaborate effectively with AI systems.

- **Human-Centered Design**: AI systems must be designed to *empower* human experts, not to marginalize them. *Hybrid Intelligence* models (Puerta-Beldarrain et al. 2025) and localized frameworks, such as Qatar's focus on participatory smart cities and sector-specific workforce development, demonstrate how context-aware strategies can preserve agency while driving innovation.

- **Appropriate Trust**: Building and maintaining trust in AI systems is crucial for effective collaboration, requiring transparency, explainability, and iterative feedback (Mehrotra et al. 2024).

- **Robust Governance**: Proactive regulations, international cooperation, and ethical guidelines are essential to mitigate systemic risks and ensure AI benefits all of humanity. Regional initiatives, like Qatar's regulatory sandboxes and HITL mandates, exemplify how governance can balance innovation with accountability.

- **Addressing the AI Governance Paradox**: Accelerating regulatory and technological oversight tools is critical to keep pace with AI advancements (ITU 2024).

By embracing these principles—and tailoring them to regional contexts, as illustrated by Qatar's alignment of AI policies with its *National Vision 2030*—we can navigate the shifting sands of the AI landscape. This approach fosters a future where human expertise is not only preserved but *enhanced* by AI, leading to a more prosperous, equitable, and empowered society. Further research should focus on validating *practical strategies* for achieving this vision, with emphasis on preventing gradual disempowerment, fostering cross-sector collaboration, and ensuring human agency remains central in an AI-driven world.

10. References

1. Alowais, S. A., S. S. Alghamdi, N. Alsuhebany, and others. "Revolutionizing Healthcare: The Role of Artificial Intelligence in Clinical Practice." BMC Medical Education 23 (2023): 689. https://doi.org/10.1186/s12909-023-04698-z.

2. Boussioux, Léonard, Jacqueline N. Lane, Miaomiao Zhang, Vladimir Jacimovic, and Karim R. Lakhani. "The Crowdless Future? Generative AI and Creative Problem-Solving." *Organization Science* 35, no. 5 (September 1, 2024): 1589–1607. https://doi.org/10.1287/orsc.2023.18430.

3. Christman. 2024. "Obtain Insights Into Complex Legal Issues With Legal Ai Summarization." LexisNexis. April 25, 2024. https://www.lexisnexis.com/community/insights/legal/b/product-features/posts/obtain-fast-insights-into-complex-legal-issues-with-legal-ai-summarization-tool?srsltid=AfmBOoq2z--msur_WaE_UXSW9c6_IWiVgkQTvmB9rYaBvrUihWPDNJ1Y.

4. Costa, Emanuele, and Nadine Moore. 2025. "GenAI Will Write the Future of Insurance Claims." BCG Global. February 6, 2025. https://www.bcg.com/publications/2023/the-future-of-insurance-claims.

5. Credo.ai. n.d. "Human-on-the-loop." https://www.credo.ai/glossary/human-on-the-

loop#:~:text=Humans%2Don%2Dthe%2DLoop,and%20intervention%20to%20continue%20improving.

6. Dell'Acqua, Fabrizio, Edward McFowland III, Ethan R. Mollick, Hila Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R. Lakhani. "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality." Harvard Business School Technology & Operations Management Unit Working Paper No. 24-013, The Wharton School Research Paper, September 15, 2023. Available at SSRN. https://ssrn.com/abstract=4573321 or http://dx.doi.org/10.2139/ssrn.4573321.

7. Felin, Teppo, and Matthias Holweg. "Theory Is All You Need: AI, Human Cognition, and Causal Reasoning." February 24, 2024. Available at SSRN. https://ssrn.com/abstract=4737265 or http://dx.doi.org/10.2139/ssrn.4737265.

8. Guo, Q., M. Ren, S. Wu, Y. Sun, J. Wang, Q. Wang, Y. Ma, X. Song, and Y. Chen. "Applications of Artificial Intelligence in the Field of Air Pollution: A Bibliometric Analysis." Frontiers in Public Health 10 (2022): 933665. https://doi.org/10.3389/fpubh.2022.933665.

9. International Telecommunication Union (ITU). 2024. "AI Governance Day - From Principles to Implementation." ITUPublications. https://s41721.pcdn.co/wp-content/uploads/2021/06/2401225_AI_Governance_Day_2024_Report-E.pdf.

10. International Trade Administration. 2024. "Qatar Artificial Intelligence." International Trade Administration | Trade.Gov. September 10, 2024. https://www.trade.gov/market-intelligence/qatar-artificial-intelligence?utm_source=chatgpt.com.

11. Ivcevic, Zorana, and Mike Grandinetti. "Artificial Intelligence as a Tool for Creativity." Journal of Creativity 34, no. 2 (2024): 100079. https://doi.org/10.1016/j.yjoc.2024.100079.

12. Kulveit, Jan, Raymond Douglas, Nora Ammann, Deger Turan, David Krueger, and David Duvenaud. "Gradual Disempowerment: Systemic Existential Risks from Incremental AI Development." arXiv preprint, 2025. https://arxiv.org/abs/2501.16946.

13. Lin, Jessy, Nicholas Tomlin, Jacob Andreas, and Jason Eisner. "Decision-Oriented Dialogue for Human-AI Collaboration." Transactions of the Association for Computational Linguistics 12 (2024): 892–911. Cambridge, MA: MIT Press. https://doi.org/10.1162/tacl_a_00679.

14. Maathuis, Clara. "Human-Centered AI in Military Cyber Operations." International Conference on Cyber Warfare and Security (ICCWS), 19, no. 1 (2024). https://doi.org/10.34190/iccws.19.1.1972.

15. Mayer, Hannah, Lareina Yee, Michael Chui, and Roger Roberts. 2025. "Superagency in the Workplace: Empowering People to Unlock AI&Rsquo;S Full Potential." McKinsey & Company. January 28, 2025. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/superagency-in-the-workplace-empowering-people-to-unlock-ais-full-potential-at-work.

16. Mehrotra, Siddharth, Chadha Degachi, Oleksandra Vereschak, Catholijn M. Jonker, and Myrthe L. Tielman. "A Systematic Review on Fostering Appropriate Trust in Human-AI

Interaction: Trends, Opportunities and Challenges." *ACM Journal of Responsible Computing* 1, no. 4 (December 2024): Article 26, 45 pages. https://doi.org/10.1145/3696449.

17. Ministry of Communications and Information Technology (MCIT). 2024. "Artificial Intelligence in Qatar – Principles and Guidelines for Ethical Development and Deployment." Ministry of Communications and Information Technology. https://prod16-assets.sprinklr.com/prod16-cdata/DAM/160136/33fb390b-d120-47f2-8708-532dcb1e15c3-1252740832/MCIT_National_AI_Strategy_-_AI.pdf.

18. Mitchell, Margaret, Avijit Ghosh, Alexandra Sasha Luccioni, and Giada Pistilli. "Fully Autonomous AI Agents Should Not Be Developed." arXiv preprint, 2025. https://arxiv.org/abs/2502.02649.

19. Mosqueira-Rey, E., E. Hernández-Pereira, D. Alonso-Ríos, and others. "Human-in-the-Loop Machine Learning: A State of the Art." Artificial Intelligence Review 56 (2023): 3005–3054. https://doi.org/10.1007/s10462-022-10246-w.

20. National Cyber Security Agency (NCSA). 2024. "Guidelines for Secure Adoption and Usage of Artificial Intelligence." National Cyber Security Agency. https://assurance.ncsa.gov.qa/sites/default/files/publications/policy/2024/CSSP_Guidelines_for_Secure_Usage_and_Adoption_of_Artificial_intelligence-Eng-v1.0_2.pdf?csrt=6420330814290588976.

21. Nguyen, Andy, Yvonne Hong, Belle Dang, and Xiaoshan Huang. 2024. "Human-AI Collaboration Patterns in AI-Assisted Academic Writing." Studies in Higher Education 49 (5): 847–64. https://doi.org/10.1080/03075079.2024.2323593.

22. Pew Research Center. "The Future of Human Agency." February 24, 2023. https://www.pewresearch.org.

23. Puerta-Beldarrain, M., O. Gómez-Carmona, R. Sánchez-Corcuera, D. Casado-Mansilla, D. López-de-Ipiña, and L. Chen. "A Multifaceted Vision of the Human-AI Collaboration: A Comprehensive Review." *IEEE Access*. https://doi.org/10.1109/ACCESS.2025.3536095.

24. Putty, Cameron. 2025. "The Power of AI in Streamlining Healthcare Workflows." Thoughtful (blog). January 10, 2025. https://www.thoughtful.ai/blog/the-power-of-ai-in-streamlining-healthcare-workflows.

25. QATAR CENTER FOR ARTIFICIAL INTELLIGENCE (QCAI). 2019. "National Artificial Intelligence Strategy for Qatar." QATAR CENTER FOR ARTIFICIAL INTELLIGENCE (QCAI). https://qcai-blog.qcri.org/wp-content/uploads/2020/04/QCRI-Artificial-Intelligence-Strategy-2019-ENG.pdf.

26. Sheharyar. 2024. "The Role of Artificial Intelligence in Financial Risk Management: Saudi Perspectives." Insights (blog). January 23, 2024. https://insightss.co/blogs/role-of-artificial-intelligence-in-financial-risk/#:~:text=AI%20Models%20for%20Risk%20Prediction,-Artificial%20Intelligence%20models&text=Machine%20learning%20models%20specifically%20designed,creditworthiness%20of%20individuals%20and%20entities.

27. The Peninsula. 2025. "Generative AI offers transformative pathway for innovation in

Qatar." February 17, 2025. https://thepeninsulaqatar.com/article/17/02/2025/generative-ai-offers-transformative-pathway-for-innovation-in-qatar.

28. Uuk, Risto, Carlos Ignacio Gutierrez, Daniel Guppy, Lode Lauwaert, Atoosa Kasirzadeh, Lucia Velasco, Peter Slattery, and Carina Prunkl. "A Taxonomy of Systemic Risks from General-Purpose AI." November 22, 2024. Available at SSRN. https://ssrn.com/abstract=5030173 or http://dx.doi.org/10.2139/ssrn.5030173.

29. Vaccaro, Michelle, Abdullah Almaatouq, and Thomas Malone. "When Combinations of Humans and AI Are Useful: A Systematic Review and Meta-Analysis." *Nature Human Behaviour* 8, no. 12 (December 1, 2024): 2293–2303. https://doi.org/10.1038/s41562-024-02024-1.