STRIKING A BALANCE

AN INTERVIEW WITH AN AI EXPERT

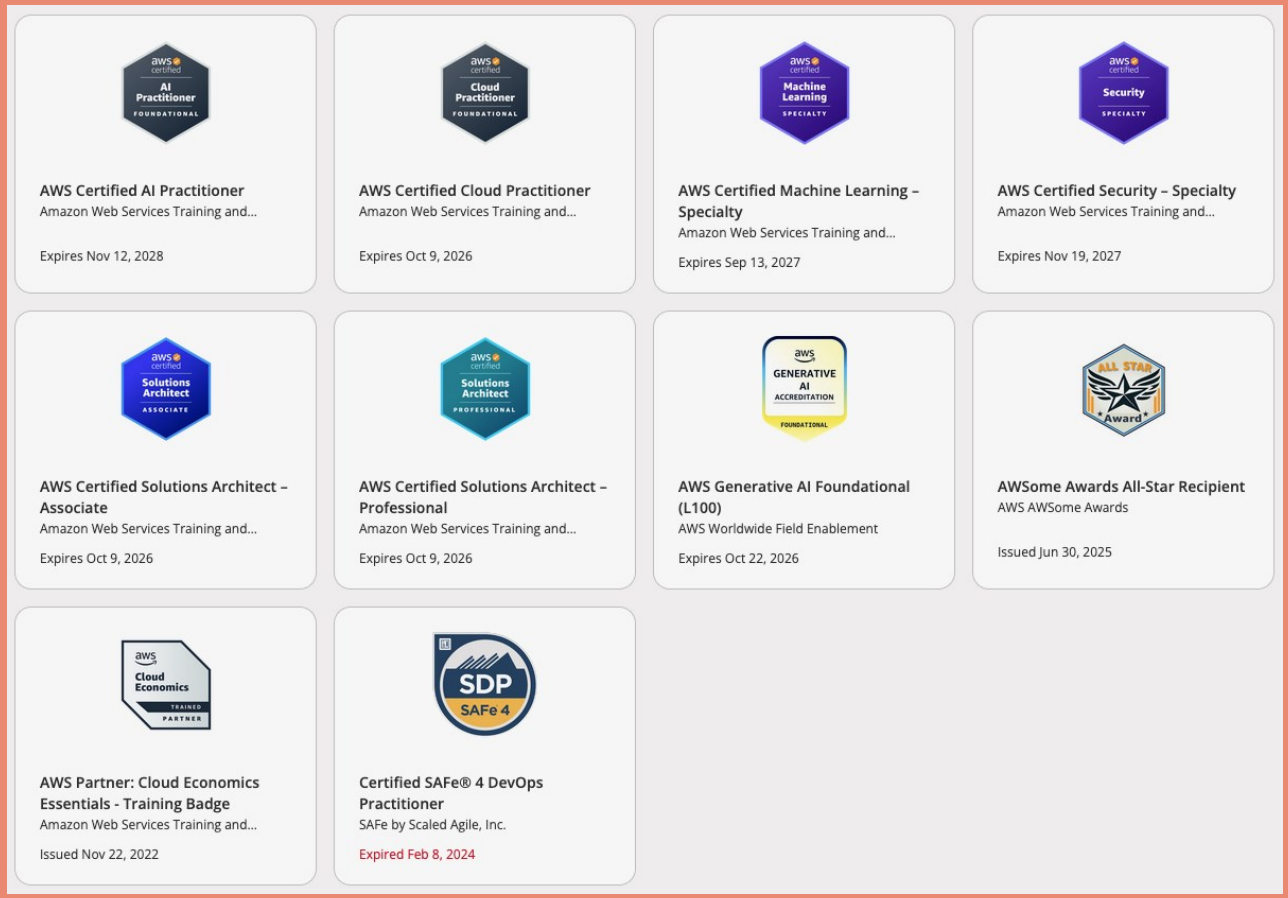
MSC3110 Final

By Joe Andaloro

INTERVIEWEE BACKGROUND

Greg Sligh

Amazon Web Services Senior Solutions Architect



AI IN MODERN COMMUNICATIONS

HOW DO YOU DEFINE AI'S ROLE?

Greg Sligh:

"AI has shifted communications from a purely human activity to a collaborative process where AI handles routine tasks, language processing, and initial content generation (drafts), while humans provide strategic direction, creative insight, and quality control."

"AI has fundamentally transformed modern communications, accelerating and enhancing human interaction across mediums."

Takeaways:

- AI now acts as a collaborative partner, managing repetitive tasks, language processing, and draft creation
- Humans focus on guiding strategy, ensuring creativity, and validating accuracy
- Oversight is essential to prevent AI from producing plausible but incorrect outputs
- Generative AI enables faster creation of high-quality content
- Chatbot performance is enhanced through careful prompt design and the use of safety or accuracy guardrails

DEMOCRATIZING EXPERTISE & INDUSTRY REVITALIZATION

DO YOU SEE AI TAKING OVER MENIAL TASKS OR COMPLETELY CHANGING WORK?

Greg Sligh:

"AI is absolutely crushing repetitive, mind-numbing tasks. But what's really fascinating is how it's not just replacing work, it's creating entirely new ways of doing business that weren't even possible before."

"A small business owner can now access marketing insights that used to require a team of analysts, or a startup can build sophisticated customer service capabilities that rival Fortune 500 companies."

"The real game-changer is that AI is creating whole new job categories and business models."

Takeaways:

- AI is automating repetitive, low-value tasks
- Beyond automation, AI is enabling entirely new business models and methods that were previously unattainable
- It supports innovations like scalable personalized services in healthcare and mass customization in manufacturing
- AI democratizes access to advanced capabilities, allowing small businesses and startups to perform tasks that once required large teams
- These developments are generating new roles and expanding opportunities across industries

VALUE OF HUMAN-IN-THE-LOOP

WHERE DO YOU SEE HUMANS WORKING IN TANDEM WITH AI IN HIGH-RISK CONTEXTS?

Greg Sligh:

"Human-in-the-loop adds the most value and is necessary in high-risk contexts where the consequences of errors are severe and irreversible."

"The greatest value is in contexts where AI provides speed and pattern recognition while humans contribute judgment, accountability, and adaptability."

"Emergency response systems use AI for rapid threat assessment and resource allocation, but human commanders make critical decisions."

Takeaways:

- Human-in-the-loop is essential in situations where mistakes carry severe or irreversible consequences
- AI provides speed and identifies patterns, while humans supply judgment, responsibility, and flexibility
- In content moderation, AI can flag potential issues, but humans evaluate context and cultural nuance
- During emergencies, AI can quickly assess threats and allocate resources, yet humans make final strategic and coordination decisions

ONE-WAY DOORS

AT WHAT POINT DOES REMOVING HUMAN OVERSIGHT BECOME UNACCEPTABLE?

Greg Sligh:

"Removing human oversight becomes unacceptable when dealing with ‘one-way door’ decisions—those that are costly or impossible to reverse."

"A 90 percent accurate recommendation engine for entertainment content is excellent, but that same accuracy level in drug dosage calculations or aircraft maintenance is catastrophically insufficient."

"Certain decisions will always require human accountability and the ability to consider factors beyond the AI's training data."

Takeaways:

- Human oversight is mandatory for decisions that are irreversible or have high stakes
- Low-risk, reversible ("two-way door") decisions can tolerate more AI autonomy
- Even highly accurate AI is insufficient for critical applications like healthcare or aviation
- Some decisions will always require human accountability to manage ethical, legal, or organizational responsibilities

AUTOMATION & HUMAN JUDGMENT

WHAT DISTINGUISHES WORK USING EITHER METHOD?

Greg Sligh:

"Traditional automation worked well for deterministic processes with predictable inputs, outputs, and well-established pathways… But they broke down when encountering unexpected data or requiring adaptive decision-making."

"The key differentiator now isn't whether a workflow can be automated, but whether it should be based on risk tolerance."

"Workflows with low stakes can operate with minimal human intervention, while those affecting safety, legality, or financials require human checkpoints."

Takeaways:

- Traditional automation is effective for predictable, rule-based tasks but struggles with unexpected or complex situations
- AI agents extend automation to complex workflows using reasoning, planning, and iterative problem-solving
- Decisions on automation should consider risk tolerance, compliance requirements, and acceptable error levels
- Low-risk, reversible workflows can run with minimal oversight, while high-stakes processes need human intervention

FAILURES

WHAT FAILURES ARE PROMINENT WHEN AI OVERFIT OR UNDER-GENERALIZE THEIR TASKS?

Greg Sligh:

"Overfitting occurs when models are trained on limited datasets and perform well on familiar scenarios but fail catastrophically when encountering new situations."

"Under-generalization happens when training data lacks sufficient variety or representation, causing models to miss critical edge cases or demographic groups."

"Beyond accuracy issues, these failures often manifest as hallucinations—where AI generates plausible but incorrect information—and social bias, where models perpetuate or amplify existing inequities in their training data."

Takeaways:

- Overfitting happens when models excel on familiar data but fail with new or unexpected inputs
- Under-generalization occurs when training data is insufficiently diverse, leading to missed edge cases or demographic groups
- These shortcomings can result in AI producing plausible but incorrect outputs (hallucinations) and reinforcing existing social biases

GUARDRAILS

HOW DO YOU AUDIT AI FOR COMPLIANCE, BIAS, AND RELIABILITY?

Greg Sligh:

"Auditing AI agents for compliance, bias, and reliability requires a multi-layered approach combining automated evaluation tools with governance frameworks."

"Comprehensive auditing involves explainability tools that provide transparency into decision-making processes, allowing organizations to understand why specific outputs were generated."

Takeaways:

- Auditing AI effectively requires multiple layers, combining automated tools with structured governance
- Real-time compliance tools can prevent policy violations before they occur
- Advanced models can continuously evaluate, rank, and validate outputs from deployed AI agents
- Explainability tools help organizations understand how and why AI decisions are made

SECURITY MISCONCEPTIONS

WHAT DO PROFESSIONALS CURRENTLY MISUNDERSTAND ABOUT AI?

Greg Sligh:

"The biggest misconception is that AI is too bleeding edge and there are too many security concerns to go into production with solutions."

"Security concerns around AI are absolutely legitimate, but they are NOT insurmountable barriers requiring perfect solutions before any adoption. ‘Wait and see’ may translate to ‘left behind.’"

"Another common misconception is that AI security is a one-time implementation… AI security requires the same continuous monitoring, updating, and skill development as any other technology stack."

Takeaways:

- A common misconception is that AI is too experimental or risky for production, causing organizations to delay adoption and fall behind
- Many also mistakenly treat AI security as a one-time setup rather than an ongoing process
- Effective AI security demands continuous monitoring, updates, and adaptive frameworks, treating risk management as an ongoing responsibility

EMERGING ATTACKS

HOW CAN AI'S WEAKNESSES BE EXPLOITED?

Greg Sligh:

"Key emerging attacks include prompt injection, where malicious inputs trick Large Language Models into disclosing sensitive information, gaining unauthorized access to connected tools, or executing commands in integrated systems."

"Other evolving threats include data poisoning during training, model extraction attempts to steal proprietary AI capabilities, and adversarial inputs designed to cause misclassification in critical applications."

Takeaways:

- Prompt injection attacks manipulate AI inputs to make models reveal sensitive information or execute unauthorized actions
- Unbounded consumption attacks overload AI systems with excessive requests, causing service disruption and financial strain
- Additional risks include data poisoning during training and attempts to extract proprietary model capabilities

FUTURE WORK ENVIRONMENTS

WHAT SKILLS DO EMPLOYEES NEED IN HYBRID HUMAN-AI WORK?

Greg Sligh:

"AI is proving the value of liberal arts education in ways we never expected."

"The ability to think critically, communicate nuanced ideas, understand context and ethical implications, and synthesize information across disciplines—these are exactly what liberal arts education develops and what hybrid AI-human environments desperately need."

"Employees need the critical evaluation skills to assess AI outputs for accuracy, bias, and appropriateness—essentially, the ability to think about meaning and consequences rather than just execution."

Takeaways:

- As AI automates routine tasks, distinctly human skills become increasingly valuable
- Critical thinking, clear communication, ethical reasoning, and interdisciplinary synthesis are essential in hybrid human-AI work
- Effective prompt engineering relies on narrative, context, and persuasive clarity to guide AI output
- Employees must evaluate AI results for accuracy, bias, and relevance, considering the broader implications of its outputs

GOVERNANCE

WHAT LONG-TERM USE OF AI DO YOU EXPECT TO SEE ACROSS INDUSTRIES?

Greg Sligh:

"I expect governance to evolve on multiple fronts simultaneously. Technical standards like Model Context Protocol (MCP) and Agent-to-Agent (A2A) communication are already enabling interoperability."

"We're also likely to see new organizational roles and governance structures. Chief AI Officers or AI Ethics Boards will become as common as Chief Security Officers are today."

"The governance frameworks that succeed will be those that enable innovation while maintaining accountability."

Takeaways:

- AI governance will develop along multiple fronts, including technical standards that support interoperability between systems
- Industry-specific regulatory frameworks, like HIPAA for healthcare, will become more prevalent
- New roles, such as Chief AI Officers and AI Ethics Boards, are likely to become standard within organizations
- Effective governance will combine audit trails and liability structures to ensure accountability for decisions made by autonomous AI systems

CONCLUSION

- AI transforms work into a collaborative process, automating routine tasks while humans provide guidance, creativity, and oversight
- Human judgment remains essential for high-stakes, irreversible, or complex decisions, ensuring accountability and ethical compliance
- AI enables new business models, democratizes expertise, and creates novel roles and opportunities across industries
- Limitations such as overfitting, under-generalization, hallucinations, and bias require continuous evaluation, auditing, and human intervention
- Security and governance are ongoing responsibilities; adaptive frameworks, standards, and transparency are critical for safe AI deployment
- Hybrid human-AI skills like critical thinking, communication, ethical reasoning, and prompt engineering are increasingly valuable
- Practical AI use such as coding or enhancing visuals demonstrates efficiency gains but still requires human refinement for quality and alignment
- In conclusion, the overall impact of AI is situational: it can accelerate creation, expand capabilities, and reduce repetitive work, but oversight, evaluation, and ethical considerations remain central

THANK YOU FOR READING!

SPECIAL THANKS TO GREG SLIGH FOR THE INFORMATION & INTERVIEW!

MSC3110 Final

By Joe Andaloro