

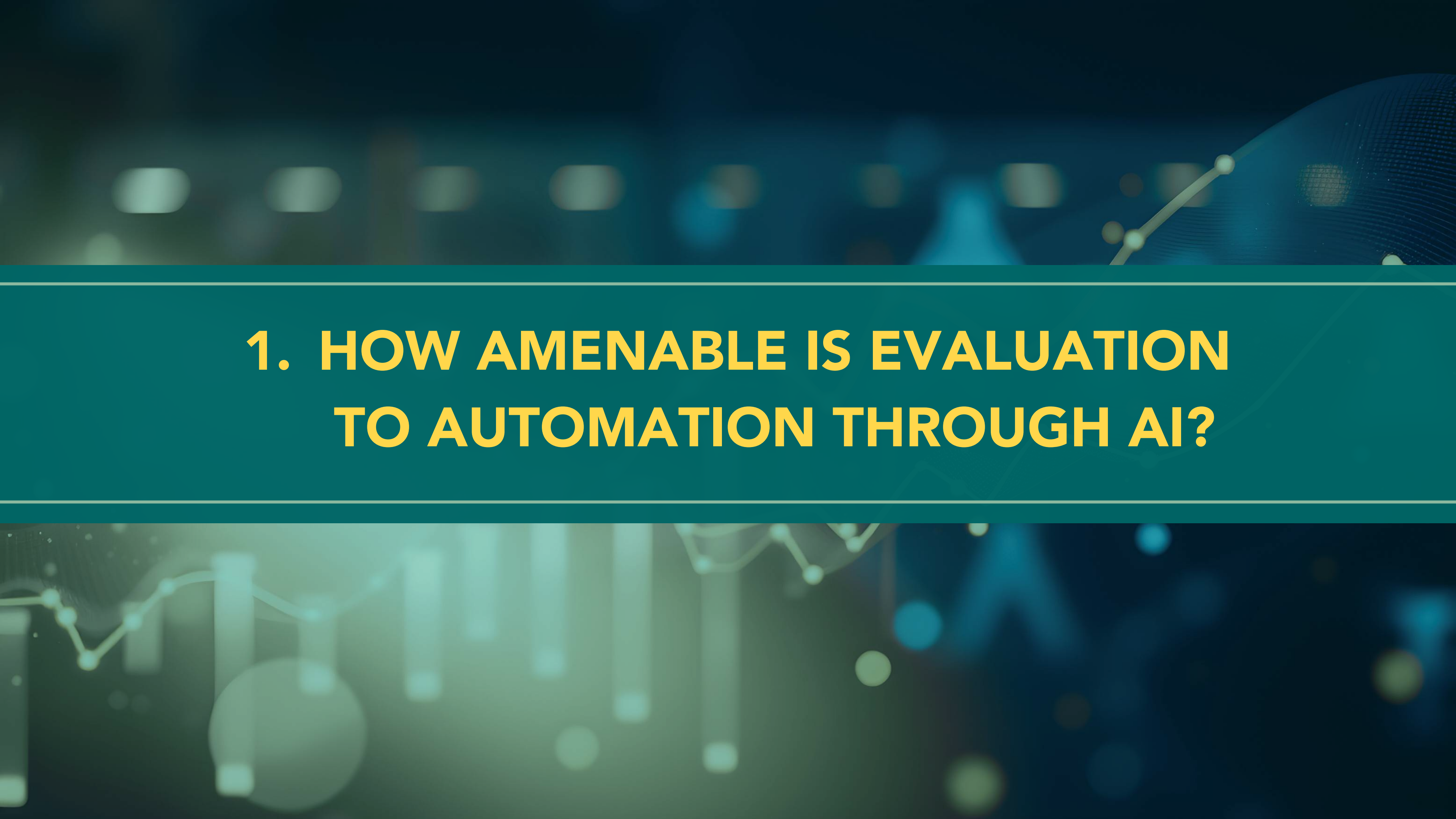


HOW WILL AI IMPACT PROGRAM EVALUATION?

Webinar, November 25th, 2025

DISCUSSION QUESTIONS

- 1 How amenable is Evaluation to automation through AI?
- 2 How will AI impact the profession?
- 3 How should you respond?
- 4 What are the technology options?

The background features a dark blue gradient with abstract data visualizations. A prominent teal horizontal band spans the middle of the image, containing the main text. Above and below this band, there are faint, stylized representations of data series, including line graphs with circular markers and vertical bar charts, all in shades of blue and teal.

1. HOW AMENABLE IS EVALUATION TO AUTOMATION THROUGH AI?

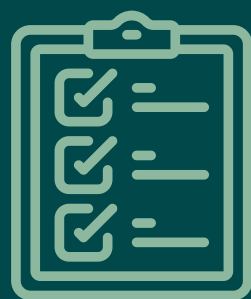
PROGRAM EVALUATION IS HIGHLY VULNERABLE TO DISRUPTION BY AI (40% TO 50% COULD BE AUTOMATED)

- Well-established knowledge base (repositories)
- Data intensive (primary and secondary)
- Standardized methodologies and processes
- Text heavy inputs and outputs (deliverables)
- Knowledgeable practitioners
- Repeatable processes/workflows
- High ease of verification vs. level of effort (asymmetries)





AI CAN BE APPLIED TO VIRTUALLY THE ENTIRE PROCESS



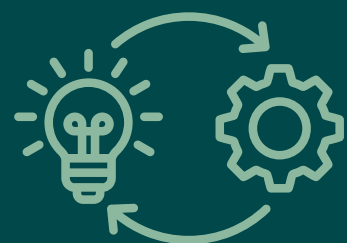
Planning (Evaluation Frameworks and Work Plans):

- Program profiles
- Logic models/Theory of Change
- Matrices (evaluation questions, indicators, and data sources)
- Methodologies
- Data fields/questionnaires and instruments



Reporting:

- Technical reports
- Final reports
- Idea generation
- Visualization
- Presentations
- Dashboards



Implementation:

- Literature review
- Environmental scans
- Survey and interview administration/ transcriptions
- Data processing/qualitative and quantitative analysis/chatting
- QA



2. HOW WILL AI IMPACT THE PROFESSION?

ADOPTION WILL ACCELERATE OVER THE NEXT 24 MONTHS.....

- To date, there has been **broad** (most evaluators are using it some) **but not deep adoption** of AI (i.e. extensive use of AI tools across a range of functions).
- However, **client demands and increasing competition** will soon change that.
- In the next 24 months, **we can expect to see**
 - Clear differentiation in terms of how effectively evaluators leverage AI
 - Increased internal evaluation capacity within users of evaluation services
 - Emergence of AI-first evaluation startups (non-traditional business models)
 - Larger consulting firms coming downstream
 - Platforms/Software-as-a-Service firms selling tools and evaluation solutions



..... RESULTING IN SIGNIFICANT CHANGES IN EVALUATION

- Rising “client” expectations (speed, breadth & depth, new methodologies and deliverables, focus on implementation)
- Evaluation capacity will increase (supply of services)
- Demand will also grow (Jevons Paradox) but likely slower than supply
- Increased commoditization/productization of traditional services
- For contractors, there will be margin compression
- Push to find new ways of differentiation (value-added/enhanced services)
- Eventual re-engineering of methodologies, outputs, and processes to leverage AI capabilities
- Redefinition of consulting roles and skill requirements
- Reduced flow of people into the profession (entry-level hiring)
- Exit of lagging professionals and firms





3. HOW SHOULD YOU RESPOND?

FIRST, RECOGNIZE THAT AI IS COMING AND IT WILL BE DISRUPTIVE

- The challenge is to advance yourself and your organization along the AI Adoption Continuum:
 - Laggards (No AI integration)
 - Experimenters (e.g., using ChatGPT)
 - Integrators (Integrating AI in formal workflows)
 - Innovators (Re-engineered processes and services to lever AI)
- Some actions you can take:
 - Review and prioritize your use cases (based on frequency, intensity, verification, data availability/readiness, governance, cost, feasibility etc.)
 - Adopt an AI Strategy and AI Governance Strategy
 - Develop a road map
 - Implement priority use cases
 - Track and evaluate the results, refine your strategy/niche and extend use cases

SECOND, RECOGNIZE EFFECTIVE AI USE INVOLVES MUCH MORE THAN JUST TECHNOLOGY

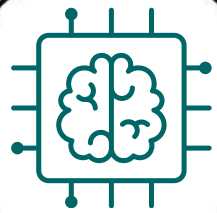
Tools and technology (LLM, scaffolding, RAG, agents, etc.) are only part of picture. You also need:

- Effective prompt and context engineering
- Data readiness (e.g., repositories of frameworks, guides, methodologies, and reports)
- Well defined roles for the Humans-in-the-Loop
 - Expert leadership (methodological expertise and professional judgement)
 - AI literacy among staff
- Validation (audit trails and QA)
- Governance (AI Usage Policies, Guidelines, Bias, Privacy, Security, Transparency, etc.)



4. WHAT ARE TECHNOLOGY OPTIONS (TIERS)?

AI APPS CAN BE GROUPED INTO 4 TIERS. PART OF YOUR STRATEGY IS DECIDING WHERE YOU FIT



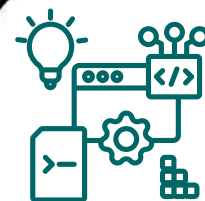
TIER 1 - DIRECT LLM INTERFACES



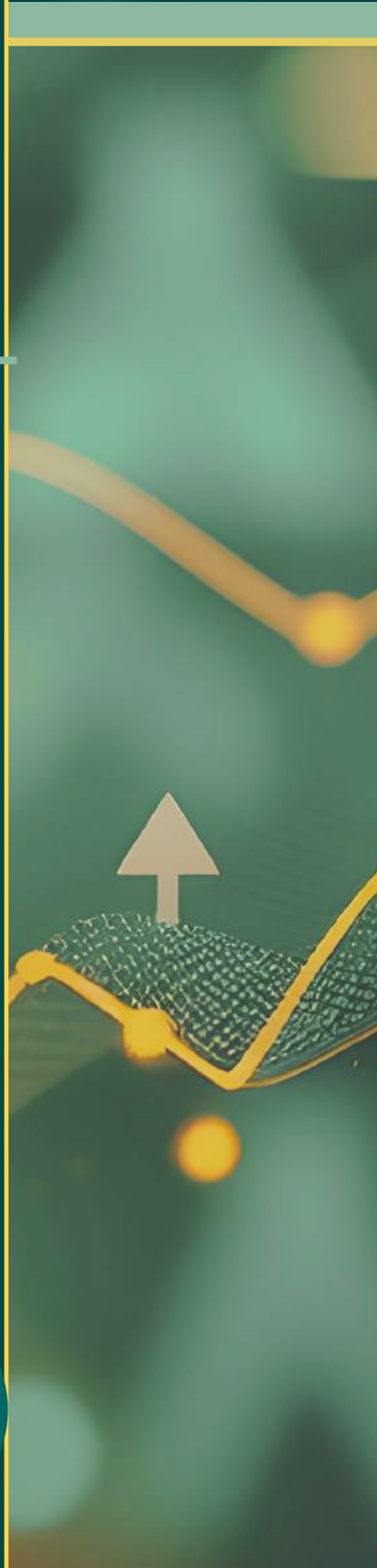
TIER 2 - COMMERCIAL AI-POWER APPLICATIONS



TIER 3 - AI ORCHESTRATION FRAMEWORKS & SCAFFOLDING



TIER 4 - FULLY CUSTOM DEVELOPMENT



TIER 1 – DIRECT LLM INTERFACES

- **Description:** Using LLMs (ChatGPT, Claude, Gemini, or Perplexity) directly through consumer-friendly web interfaces or APIs.
- **Advantages:** Accessibility, no technical setup required, rapid experimentation, low cost, simple to learn, no vendor lock-in
- **Disadvantages:** Concerns regarding data security/privacy, process & system integration/customization, repeatability, verification (no audit trails), scalability & standardization, governance, context, access to your data
- **Best For:** experimenting with AI, quick ad-hoc tasks, exploring possibilities, testing ideas

TIER 2 – COMMERCIAL AI-POWERED APPS

- **Description:** Established apps that integrated AI (e.g. NVivo, MAXQDA, Qualtrics, SurveyMonkey with AI, etc.) and AI-first apps (Outset, CoLoop, Leading AI, Otter.ai, Insight7, etc.)
- **Advantages:** Purpose-built (for specific functions), enterprise grade, team-friendly, audit trails and governance, data remains in-system, vendor support and training, regular updates, budget predictability
- **Disadvantages:** Higher cost, very-function specific/limited integration, narrow range of functions, limited customization, vendor dependence, feature lag, one-size-fits-all, governance and scalability constraints, differentiation
- **Best For:** Firms with sufficient volume of targeted, isolated tasks, that prioritize ease of use and compliance, and have limited technical capacity for custom development

TIER 3 - AI ORCHESTRATION FRAMEWORKS & SCAFFOLDING

- **Description:** Connects LLMs to your unique data, workflows, and methodologies through orchestration platforms, scaffolding layers and RAG.
- **Advantages:** Differentiation, proprietary knowledge integration, automation of customizable workflows, embedded learning, context and consistency, data access/control/privacy, scalability, programming of audit trails/validation processes, integration across functions
- **Disadvantages:** Moderate technical skill required, implementation time, ongoing maintenance, development complexity, team capability building, governance development required, and cost variability.
- **Best For:** Firms with some technical capability and significant proprietary methodologies and data, wanting to differentiate from competitors

TIER 4 - FULLY CUSTOM DEVELOPMENT

- **Description:** Entire systems including data architecture, user interfaces, custom dashboards, AI models, and proprietary algorithms. This might include specialized evaluation platforms, proprietary analysis engines, or productized solutions you market to other evaluators.
- **Advantages:** Maximum customization, proprietary technology, full data ownership and control, IP protection, scaling advantages, productization potential, complete integration, and team differentiation,
- **Disadvantages:** Highest cost and time investment, technical leadership required, long development timeline, team skill requirements, maintenance burden, scaling complexity, AI model management, market timing risk, governance responsibility, staff dependency, opportunity cost,
- **Best For:** Larger firms with substantial resources, technical capability, highly specialized methodologies, and maybe planning to productize and sell evaluation tools

MOST EVALUATORS WILL USE MULTIPLE TIERS

Tier 1 for quick exploration,
Tier 2 for standard workflows,
and Tier 3 or 4 (directly or through partnerships) for competitive differentiation

FACTOR		1 Direct LLM	2 Commercial Apps	3 Orchestration	4 Custom
\$	COST	Low	Medium	Medium-High	Very High
🕒	TIME TO VALUE	Days	Days-Weeks	Weeks-Months	Months-Years
🧑‍🔧	TECHNICAL SKILLS REQUIRED	None	Minimal	Moderate	High
🔒	DATA SECURITY	Low	Medium-High	High	Very High
⚙️	CUSTOMIZATION	None	Limited	Extensive	Complete
🌐	SCALABILITY	Limited	Good	Excellent	Excellent
🔄	ORGANIZATIONAL LEARNING	Minimal	Low	High	Very High
🏆	BEST FIT	Individuals/ Freelancers	Small/Medium Teams	Firms with proprietary methods	Larger orgs with technical expertise