Lesson 2

Advancing AI Fluency

AINOM® Course - Attending this course gives learners access to related course materials.

AINOM
PROVIDED BY SCALED AGILE

Learning Objectives

By the end of this lesson, you will:

- Explain how AI Fluency helps Change Agents guide AI Solutions to success while avoiding common traps
- Translate technical AI concepts like ML, DL, and GenAI into simple business language
- Review LLMs, RAG, and Agentic solutions to strengthen your AI understanding
- Evaluate your organization's existing AI tools and show how reusing them saves time and money
- Help teams pick realistic AI approaches instead of chasing impossible dreams

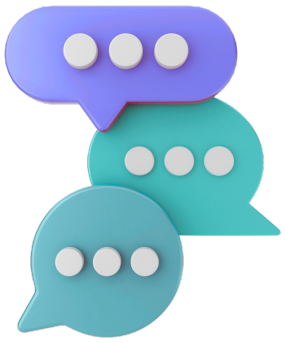
WIIFM: Gain the confidence to guide technical teams, ask smarter questions, and avoid unrealistic AI hype.

Translation Problems

5 min

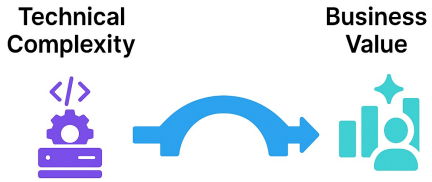
Group Discussion:

- When technical teams and business teams try to work together on initiatives, what usually gets lost in translation?
- Share specific examples.



Why AI Fluency Matters

- **AI Creates New Challenges:**
  - **Traditional software**: Predictable, rule-based, deterministic outputs
  - **AI systems**: Probabilistic, pattern-based, variable outputs
  - **Result**: New failure modes, costs, and risks teams don't anticipate
- **Your Role as an AI-Native Change Agent:**
  - Bridge technical complexity and business value
  - Spot feasibility traps before resources are wasted
  - Guide teams to solutions that ship and realize value.
  - Prevent the 80% failure rate
  - Your role isn't to be the AI expert, but to help everyone make informed decisions

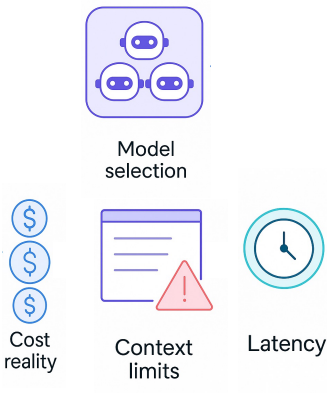


AI Fluency Concepts for AI-Native Change Agents

1. Building on LLM Foundations
2. Technical Concepts That Drive Decisions
3. Fine-Tuning: Teaching AI Your Organization's Language
4. Retrieval-Augmented Generation: Your Organization's Memory
5. AI Agents: From Assistants to Autonomous Workers
6. Solution Pattern Comparison
7. Assessing Organizational Readiness
8. From Hype to Reality - Feasibility Filters
9. Spot Red Flags and Early Warning Signs
10. Integration and Production Realities

Building on LLM Foundations

- **You Already Know:**
  - LLMs predict the next tokens
  - They're trained on massive datasets
  - They can generate human-like text
- **Now Add Production Context:**
  - **Model Selection**: GPT-4, Claude, Gemini, and more
  - **Cost Reality**: $0.01-0.03 per 1000 tokens adds up fast
  - **Context Limits**: 128K tokens ≠ , unlimited memory
  - **Latency Issues**: Larger models = slower responses
- **REMEMBER:** Success Factor 2 - Upskill Relentlessly means your AI learning never stops



Change Agent Insight: "The best-performing model isn't always the right model for your use case."

Technical Concepts That Drive Decisions

- **Parameters (70B vs 400B)**
  - **What**: Model size and complexity
  - **Impact**: Larger = more capable but expensive
  - **Decision**: "Can we achieve goals with a smaller model?"
- **Context Windows**
  - **What**: AI's working memory limit
  - **Impact**: Constrains document size and conversation length
  - **Decision**: "How much information/context needs to be passed to the LLM?"
- **Inference Costs**
  - **What**: Price per API call
  - **Impact**: $1K/day can happen quickly at scale
  - **Decision**: "Have we modeled costs at full deployment?"
- **Response Latency**
  - **What**: Time to generate answers
  - **Impact**: User experience and throughput
  - **Decision**: "Is real-time response critical?"

Fine-Tuning: Teaching AI Your Organization's Language

- **What Fine-Tuning Does:**
  - **Specializes**: Adapts model behavior to specific domains
  - **Learns Patterns**: Internalizes your terminology and style
  - **Embeds Knowledge**: Bakes information directly into the model
  - **Customizes Output**: Matches your exact format needs
- **Fine-Tuning Complexity Levels:**
  - **Simple**: Output formatting (JSON structures)
  - **Moderate**: Industry terminology (medical coding)
  - **Complex**: Behavioral adaptation (company voice)

Where Fine-Tuning Struggles

- ✗ Updating requires full retraining
- ✗ High-quality training data requirements
- ✗ Risk of catastrophic forgetting
- ✗ Expensive computing and expertise are needed

Where Fine-Tuning Shines

- ✓ Consistent specialized terminology
- ✓ Domain-specific behavior (medical, legal)
- ✓ Unique output formats and styles
- ✓ Reduced prompt engineering needs

Retrieval-Augmented Generation: Your Organization's Memory

- **How RAG Works:**
  - **Prepare**: Convert documents to searchable vectors
  - **Retrieve**: Find relevant content for each query
  - **Augment**: Combine retrieved info with LLM
  - **Generate**: Produce an answer with citations
- **Change Agent Question:** "Is our knowledge structured enough for RAG?"
- **REMEMBER:** Success Factor 5 - Tell the Full Story

Where RAG Struggles
✗ Poor quality source documents
✗ Contradictory information
✗ Complex reasoning across sources
✗ Real-time data needs

Where RAG Shines
✓ Company policies and procedures
✓ Technical documentation
✓ Knowledge that changes frequently
✓ Compliance and audit requirements

AI Agents: From Assistants to Autonomous Workers

- **Core Agent Capabilities:**
  - **Goal Understanding**: Interprets objectives
  - **Planning**: Breaks down complex tasks
  - **Tool Use**: Integrates with systems/APIs
  - **Adaptation**: Handles unexpected situations
- **Agent Complexity Levels:**
  - **Simple**: Single-task automation (email classifier)
  - **Moderate**: Multi-step workflows (expense approval)
  - **Complex**: Autonomous decision-making (trading bot)

Where Agents Struggle
✗ Unpredictable edge cases
✗ Complex judgment calls requiring ethics/values
✗ Explaining their decision-making process
✗ Recovering from cascading errors

Where Agents Shine
✓ Repetitive multi-step workflows
✓ 24/7 availability for routine tasks
✓ Consistent execution of procedures
✓ Integrating multiple systems/tools

Solution Classification Comparison

- **The AI-Native Change Agent's Decision Framework:**
  - Start with the simplest solution that could work
  - Add complexity only when justified by value
  - Consider the total cost of ownership, not just the build cost
  - Plan for failure modes from day one
- **Success Factor 4: Start Smart** - Right-size the solution

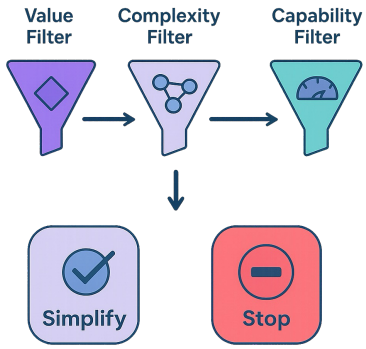
Classification	Best For	Key Risk
Prompt-Only	Creative tasks, general knowledge	No private data access
RAG	Company knowledge Q&A	Source quality dependency
Fine-Tuning	Specialized behavior	Expensive to update
Agentic	Complex automation	Reliability at scale

Assessing Organizational Readiness

- **Take Inventory Before You Build**
  - **Models**: What's already approved and available?
  - **Data**: What's clean, governed, and accessible?
  - **Infrastructure**: Cloud resources, GPUs, APIs?
  - **Expertise**: Who knows AI, data, and integration?
  - **Governance**: What are the rules and constraints?
- **Change Agent Action:** Take early opportunities to discover the organizational capabilities and policies.
- **Note:** This is not a blocker to convening the AI-Native Value Workshop

From Hype to Reality - Feasibility Filters

- **Value Filter**
  - Can we quantify the benefit?
  - Who specifically wins?
  - What's the cost of not doing it?
- **Complexity Filter**
  - What's the simplest version?
  - What could break?
  - How many moving parts?
- **Capability Filter**
  - Do we have the skills?
  - Can we maintain it?
  - Will it scale?
- If any filter shows red flags, simplify or stop
- Focus on the stable Solution, not the frontiers. Don’t be the guinea pig.



Your Mantra: "Better to ship simple than dream complex."

Spot Red Flags and Early Warning Signs

- **Technical Red Flags:**
  - "We'll just fine-tune GPT-4" *(Not possible)*
  - "RAG will solve all our problems" *(It won't)*
  - "We don't need to worry about prompts" *(You do)*
  - "The model will learn from production" *(Dangerous)*
- **Process Red Flags:**
  - No clear success metrics
  - No data quality assessment
  - No fallback plan
  - No ongoing maintenance budget
- **AI-Native Change Agent Response:**
  - Don't criticize -> educate
  - Extract simpler alternatives
  - Focus on achievable wins
  - Build credibility through small successes

Integration and Production Realities

- **Common Integration Challenges:**
  - **API Limits**: Rate limits break at scale
  - **Latency Stacking**: Each service adds delay
  - **Error Cascades**: One failure breaks everything
  - **Cost Scaling**: Dev costs scale in production
- **MLOps Requirements Often Missed:**
  - Model versioning and rollback
  - Performance monitoring
  - Data drift detection
  - A/B testing capability
  - Audit trails
- **Common Gaps That Derail Solutions:**
  - No data governance strategy
  - Underestimating integration complexity
  - Unclear ownership model
- **AI-Native Change Agent Response :**
  - "How will this work at 10x volume?"
  - "What happens during an outage?"
  - "How do we update without breaking things?"

2.1 Match the Method

15 min

Group Activity

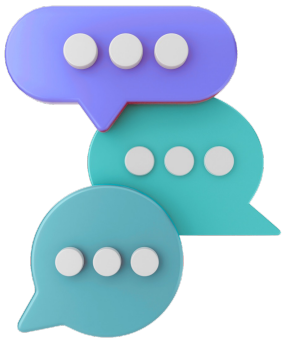
- 1. Matching
  - In the companion app, you will see four business scenarios and four AI approaches.
  - Individually, drag and drop the best AI approach—Prompt-only, RAG, Fine-tuning, or Agentic—to match each scenario.
- 2. Group Discussion
  - In your groups, compare your individual matches.
  - For each scenario, discuss and agree on the best-fit AI approach.
  - Explain the 'why' behind your choice. What key characteristic of the scenario made one approach better than the others?
- Be ready to share critical trade-offs (e.g., cost, complexity, speed, data needs) for at least two of the patterns your group discussed.

Practical Risk vs Perfect Solution

5 min

Class Discussion:

- During the activity, you probably debated some of your matches. Multiple patterns could work for any given problem, but a Change Agent's job is to find the best fit for right now.
- Questions
  - What is the single biggest risk of choosing a solution that is possible, but not the best fit?
  - How does your AI Fluency help you guide the team to the most practical starting point?



Lesson Review

You can:

- Explain how AI Fluency helps Change Agents guide projects to success while avoiding common traps
- Translate technical AI concepts like ML, DL, and GenAI into simple business language
- Review LLMs, RAG, and Agentic solutions to strengthen your AI understanding
- Evaluate your organization's existing AI tools and show how reusing them saves time and money
- Help teams pick realistic AI approaches instead of chasing impossible dreams

Insights & Action

Post-Class

Reflect

Think about a recent meeting where technical and business teams struggled to understand each other. Write about how having AI fluency might have changed that conversation. What questions could you have asked to bridge the gap?

Make notes in your journal

Apply

Ask AI: "I need to explain [insert AI concept like RAG or fine-tuning] to my non-technical manager. You're my communication coach. Help me create a simple analogy and a 2-minute explanation that avoids jargon while keeping the key points accurate."

Explore

"RAG vs. Fine-tuning and more" - Google Cloud:

"LLM Leaderboard - Model Comparisons" - Vellum.AI