90% FAIL RATE EXPECTED

Practice FTE Exam

Agent Factory Mastery Assessment

Designed by: Afnan Zahed, PIAIC, Islamabad

LinkedIn Profile | WhatsApp Group

Purpose: Practice and Prepare for the FTE Exam

Source: FTE Slides Content | Zia Khan

70	8	90%
QUESTIONS	SECTIONS	TO PASS

Questions 61–70

WARNING: If you're scoring above 85%, you're ready. Below that means you're memorizing, not understanding.

SECTION A

THE PARADIGM SHIFT

> *"If you don't understand WHY we're here, you'll never understand WHERE we're going."*

1 THE ECONOMIC THESIS

SCENARIO

Your startup just raised $10M. Your investor asks: "What's the fundamental economic shift that makes Agent Factories valuable?" You need to give a one-sentence answer that would satisfy a Stanford economics professor.

A We're moving from selling software licenses to selling AI subscriptions at scale.

B We're moving from selling tools that help humans work to selling the work itself.

C We're moving from human employees to robot employees that cost less per hour.

D We're moving from cloud computing to edge computing for faster AI inference.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A is a trap. Subscriptions are just a pricing model, not an economic shift. SaaS has existed for 20 years. This answer confuses the delivery mechanism with the fundamental value proposition.

C is dangerously wrong. "Robot employees" conflates physical robots with digital agents. The document explicitly separates Agents (digital work) from Robots (physical work). Also, "cost less per hour" is reductive—the shift is about selling outcomes, not cheaper labor.

D is irrelevant. Edge computing is an infrastructure detail, not an economic thesis. This answer would get you laughed out of a board meeting.

WHY B IS CORRECT:

The document states verbatim: "we are moving from selling tools that help humans work, to selling the work itself." This is the core thesis. Tools require humans to operate them. Digital FTEs replace the human operator entirely. You're not selling a hammer; you're selling a built house.

2 THE UI → INTENT SHIFT

SCENARIO

A Fortune 500 CTO says, "I don't get it. We spent $50M on SAP. Why would we need agents?" Which response demonstrates you understand the paradigm shift?

A "SAP is old technology. Agents use modern AI which is faster and cheaper."

B "Your employees currently act as manual routers between SAP screens. Agents eliminate that friction by translating intent directly into system actions."

C "Agents can replace SAP entirely, so you can save that $50M."

D "SAP doesn't have a chatbot interface, but agents do, so employees can ask questions in English."

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A is shallow. "Old vs. new" and "faster and cheaper" shows zero understanding of the architectural shift. This is how a salesperson talks, not an engineer.

C is dangerous. Agents don't replace ERP systems—they interface WITH them. The document never suggests throwing away your tools; it suggests eliminating the human routing layer.

D confuses chatbots with agents. The document explicitly distinguishes chatbots (passive, prompt-driven) from agents (active, goal-driven). A chatbot interface to SAP is not an agent—it's just a new UI.

WHY B IS CORRECT:

The document describes the old paradigm as users being "manual routers" who "navigate menus, click buttons, and fill out forms." The agent eliminates this by translating user INTENT directly into system actions. The CTO's SAP investment isn't wasted—but his employees' time navigating it is.

3 PROGRAMMING EVOLUTION SEQUENCE

SCENARIO

You're teaching a bootcamp. A student asks: "What's the correct historical evolution of programming as described in the Agent Factory methodology?"

A Assembly → Python → AI-Driven Development → English Prompting

B AI-Driven Development → Programming in Python → Prompting in Natural Language

C Prompting in Natural Language → Programming in Python → AI-Driven Development

D Programming in Python → AI-Driven Development → Prompting in Natural Language

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A includes Assembly which is never mentioned in the source material. This tests whether you're adding outside knowledge or sticking to the document.

C is completely backwards. Natural language prompting is the FINAL stage, not the first.

D reverses the first two stages. AI-Driven Development (autocomplete) came BEFORE we realized we could abstract further.

WHY B IS CORRECT:

The document lists exactly: "1) AI Driven Development—We started by using AI to help us autocomplete code. 2) Programming in Python—We moved to building scripts and backends explicitly. 3) Prompting in English—Now, we have reached a layer of abstraction where natural language is the compiler."

4 THE DEVELOPER'S NEW IDENTITY

SCENARIO

Your junior developer says, "I'm scared AI will take my job because it writes code faster than me." Based on the Agent Factory philosophy, what's the correct response?

A "You're right to be scared. Learn prompt engineering or become obsolete."

B "Don't worry, AI can't write good code. You'll always be needed to fix its mistakes."

C "Your job isn't to type code anymore. You're now an orchestrator—the architect who manages and audits the AI bricklayers."

D "Focus on becoming a specialist in one language. AI is generalist, but experts will always be valued."

CORRECT ANSWER **C**

WHY THE OTHERS FAIL:

A is fear-mongering without insight. It offers no strategic path forward.

B is denial. The document explicitly states that agents CAN write code, debug, and self-correct. Pretending otherwise is dangerous.

D misunderstands the shift. Specializing in syntax is exactly what becomes LESS valuable. The document says developers are no longer "typists responsible for writing every syntactic character."

WHY C IS CORRECT:

The document states: "The role has evolved into a 'developer-as-orchestrator.' Your job is to manage, guide, and audit the AI agents that write the underlying code. You are the architect; the agents are the bricklayers." This is the new professional identity.

5 THE WORKFORCE TRIAD

SCENARIO

A government official asks you to explain the future workforce structure for a policy paper. Which statement accurately represents the Agent Factory model?

A AI replaces all repetitive human jobs; humans focus only on creative work; robots are irrelevant.

B People handle judgment/ethics/edge cases; Agents handle digital automation; Robots handle physical automation.

C Agents and Robots are the same thing—both are AI systems that replace human labor.

D People supervise Agents; Agents supervise Robots; it's a strict hierarchy.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A is too extreme. The document never says AI replaces ALL repetitive jobs. It describes a partnership, not replacement. Also, robots are explicitly included.

C conflates Agents and Robots. The document explicitly separates them: Agents work with "bits" (digital), Robots work with "atoms" (physical). This distinction is fundamental.

D invents a hierarchy that doesn't exist in the source material. The document describes a collaborative ecosystem, not a command chain.

WHY B IS CORRECT:

The document explicitly lists: "People: Provide high-level judgment, creative direction, ethical oversight, and handle edge cases. Agents: Handle the high-volume, repetitive digital work automation. Robots: Handle the physical work automation, moving atoms based on digital instructions."

6 THE AI RELATIONSHIP TRIANGLE

SCENARIO

During a philosophy of AI debate, someone challenges you: "What's the fundamental nature of the human-AI relationship according to Agent Factory thinking?"

- **A** AI is a tool that humans control completely.
- **B** AI is a threat that must be carefully constrained and supervised.
- **C** AI is simultaneously teacher, student, and co-worker—the relationship is fluid.
- **D** AI is a replacement for human workers in most contexts.

CORRECT ANSWER **C**

WHY THE OTHERS FAIL:

A (tool only) misses the teaching and learning aspects. Tools don't teach you new concepts.

B (threat focus) represents a fear-based view not present in the source material's framing.

D (replacement) contradicts the "partnership" model emphasized throughout.

WHY C IS CORRECT:

The document quotes: "AI is my Teacher, my Student, and my Co-Worker. This relationship is fluid. Sometimes the AI teaches us new concepts; sometimes we teach it specific domain skills; and most of the time, we work alongside it."

7 THE VALUE COMPANY DEFINITION

SCENARIO

An MBA student asks: "What defines the most valuable technology companies of the next decade?"

A Companies with the largest user bases and network effects.

B Companies with the best LLM models and most training data.

C Companies that manufacture AI employees—autonomous digital workers at scale.

D Companies that provide the cheapest cloud computing infrastructure.

CORRECT ANSWER **C**

WHY THE OTHERS FAIL:

A (user bases) describes Web 2.0 thinking, not the Agent Factory thesis.

B (best models) confuses the ingredient with the product. Models are components; Digital FTEs are the product.

D (cheap infrastructure) is a race to the bottom, not a value creation strategy.

WHY C IS CORRECT:

The document opens with: "The most valuable companies of the next decade will not merely sell software subscriptions or seat licenses. Instead, they will manufacture AI employees."

8 WHAT AGENTS ARE NOT

SCENARIO

A journalist writes: "Agent Factory is just rebranding chatbots." How do you correct them?

A "You're partially right—agents are advanced chatbots with better prompts."

B "Agents are completely different—they use code, not natural language."

C "Agents are autonomous digital workers with goals, tools, state, and self-correction—chatbots just answer questions."

D "Agents are chatbots deployed in enterprise environments with security."

CORRECT ANSWER C

WHY THE OTHERS FAIL:

A concedes the false premise. Agents are fundamentally different, not "advanced chatbots."

B is incorrect. Agents DO use natural language—but they also use tools, maintain state, and pursue goals autonomously.

D confuses deployment context with fundamental architecture. An enterprise chatbot is still just a chatbot.

WHY C IS CORRECT:

The document explicitly distinguishes agents from chatbots across four dimensions: goal-driven (not prompt-driven), tool use (not just text generation), state/memory (not stateless), and autonomy (self-plans and self-corrects without micromanagement).

9 THE DEMOCRATIZATION CLAIM

SCENARIO

A coding bootcamp says: "You still need to learn Python to build agents." Based on the Agent Factory vision, is this true?

A Absolutely true—Python is the foundation of all agent development.

B Partially true—you need Python for Custom Agents but not General Agents.

C Increasingly false—natural language is becoming the compiler, democratizing creation beyond Python experts.

D False—agents only use JavaScript and TypeScript.

CORRECT ANSWER C

WHY THE OTHERS FAIL:

A is outdated thinking. The document explicitly challenges this view.

B makes an arbitrary distinction not supported by the source material.

D is factually wrong and easily eliminated.

WHY C IS CORRECT:

The document states: "You can program complex systems in English or Urdu, democratizing software creation beyond the elite circle of Python or TypeScript experts." This is the democratization thesis—natural language as the new programming interface.

10 THE CORE INFRASTRUCTURE

SCENARIO

You're pitching to a VC. She asks: "What are the five core technologies that power Agent Factory digital workers?"

- **A** LLMs, GPUs, Cloud Storage, APIs, and Databases
- **B** Agents, Specifications, Skills, MCP, and Cloud-Native Technologies
- **C** Python, Docker, Kubernetes, TensorFlow, and REST APIs
- **D** ChatGPT, Midjourney, GitHub Copilot, Notion AI, and Jasper

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A lists generic AI infrastructure but misses the Agent Factory-specific components.

C lists DevOps tools that are implementation details, not the conceptual framework.

D lists consumer AI products that are completely irrelevant to the architecture.

WHY B IS CORRECT:

The document explicitly states these workers are "powered by sophisticated agents, rigorous specifications (specs), modular skills, the Model Context Protocol (MCP), and cloud-native technologies."

SECTION B

THE THREE WAVES & AGENT DEFINITION

> *"Know your history or repeat its failures."*

11 WAVE CLASSIFICATION

SCENARIO

Your AI categorizes customer sentiment as "positive" or "negative" based on historical review data. Which wave does this belong to?

A Predictive AI—it's analyzing patterns to forecast sentiment.

B Generative AI—it's creating the sentiment labels.

C Agentic AI—it's autonomously processing customer data.

D None—sentiment analysis isn't AI.

CORRECT ANSWER **A**

WHY THE OTHERS FAIL:

B is wrong. Generative AI creates NEW content (text, images, code). Classification uses existing labels—it doesn't generate them.

C is a trap. Just because something processes data automatically doesn't make it "Agentic." Agentic AI is defined by autonomous goal pursuit with self-correction loops, not simple classification.

D is absurd and easily eliminated.

WHY A IS CORRECT:

Predictive AI "focused on analyzing massive historical datasets to forecast future trends" and "enabled data-driven decision-making." Sentiment classification fits this pattern—using historical labeled data to predict the sentiment of new inputs.

12 THE GENERATIVE AI LIMITATION

SCENARIO

Your marketing team loves ChatGPT for writing emails. A colleague says "ChatGPT is an agent." Why is this wrong?

A ChatGPT is too slow to be an agent.

B ChatGPT is largely passive—it waits for prompts and produces output, not autonomous action.

C ChatGPT doesn't use Python, which all agents require.

D ChatGPT was made by OpenAI, and agents can only be made by Anthropic.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A (speed) is irrelevant to the definition of an agent.

C (Python requirement) is fabricated—no such requirement exists.

D (company restriction) is obviously absurd.

WHY B IS CORRECT:

The document describes Generative AI as "still largely a 'chatbot' experience—waiting for a human prompt to produce an output." Agentic AI, by contrast, "becomes proactive" and "manages complex tasks end-to-end." ChatGPT in its basic form is reactive, not proactive.

13 THE AGENT LOOP SEQUENCE

SCENARIO

You're debugging an agent that keeps failing. To diagnose, you need to know the correct operational loop. What is the precise sequence?

A Decide → Act → Observe → Learn → Repeat

B Observe → Decide → Act → Learn → Repeat

C Learn → Observe → Decide → Act → Repeat

D Act → Learn → Observe → Decide → Repeat

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A starts with Decide. You can't decide before you observe the environment—that's reckless action.

C starts with Learn. You can't learn without having taken action and received feedback first.

D starts with Act. Acting blindly without observation is chaos, not intelligence.

WHY B IS CORRECT:

The document explicitly states the agent "Observes its environment → Decides what to do next → Takes Actions → Learns from feedback → Repeats." Observation MUST come first—you can't make intelligent decisions without data about your environment.

14 AGENT VS. CHATBOT - THE GOAL TEST

SCENARIO

System A responds to "Write a poem about cats." System B executes "Pull Xero trial balance daily, reconcile against bank feeds, export CSV." Which is the agent?

A System A—it generates creative content autonomously.

B System B—it pursues a multi-step goal with tool use across time.

C Both are agents—they both complete tasks.

D Neither is an agent—agents require physical robot bodies.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A confuses content generation with agency. Generating a poem is a single-shot response to a prompt—no goal pursuit, no tool use, no state tracking.

C dilutes the definition. Not everything that "completes a task" is an agent. A calculator completes tasks.

D invents a physical requirement that doesn't exist. Agents are explicitly digital.

WHY B IS CORRECT:

The document uses this exact example: A chatbot "answers a question ('Write a poem')" while an agent "accepts a mission ('Pull the Xero trial balance daily, reconcile it against bank feeds, and export the CSV')." The difference is goal-driven vs. prompt-driven, multi-step execution vs. single response, and tool use across time.

15 THE FOUR DEFINING CHARACTERISTICS

SCENARIO

A PM asks: "What exactly makes something an 'Agent' instead of a fancy script?" List the four characteristics.

A. Fast processing, large context window, good prompts, fine-tuning

B. Goal-driven behavior, tool use, state/memory, autonomy (self-correction)

C. Natural language input, JSON output, API integration, cloud hosting

D. Training data, model weights, inference speed, parameter count

CORRECT ANSWER B

WHY THE OTHERS FAIL:

A describes model capabilities, not agent characteristics.

C describes interface formats, not behavioral properties.

D describes LLM internals, which are implementation details, not agent definitions.

WHY B IS CORRECT:

The document explicitly lists what makes agents different: "Goal-driven vs. Prompt-driven," "Tool Use" (can call functions, query databases, browse web), "State & Memory" (knows past actions, plans future ones), and "Autonomy" (plans, executes, checks, corrects without human intervention).

16 LARGE ACTION MODELS (LAMS)

SCENARIO

Someone says "LLMs are all we need for AI." What's the Agent Factory counter-argument?

A LLMs are too expensive; LAMs are cheaper.

B LLMs understand; LAMs execute. Understanding without action is incomplete.

C LAMs have more parameters than LLMs.

D LLMs are made by OpenAI; LAMs are made by Anthropic.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A (cost comparison) isn't the fundamental distinction.

C (parameter count) is fabricated and misses the conceptual difference.

D (company attribution) is nonsense.

WHY B IS CORRECT:

The document states: "We are transitioning from Large Language Models (LLMs)—which are engines of understanding—to Large Action Models (LAMs)—which are engines of execution." Understanding without action means the AI can tell you what to do but can't do it. LAMs bridge that gap.

17 THE FIVE POWERS

SCENARIO

You're designing an agent for a warehouse. Which "power" allows it to process shipping labels visually?

A Hear—it listens to barcode beeps.

B See—visual understanding of documents and physical spaces.

C Reason—it thinks about what the label means.

D Remember—it recalls previous shipping labels.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A (Hear) is for audio processing (meetings, voice commands), not visual labels.

C (Reason) is for decision-making, not perception. Reasoning comes AFTER seeing.

D (Remember) is for long-term context, not immediate visual processing.

WHY B IS CORRECT:

The document defines See as "Visual understanding allows agents to process invoices, analyze screen UIs, and navigate physical spaces." Processing a shipping label is visual understanding—the "See" power.

18 THE "REMEMBER" POWER

SCENARIO

Your agent worked on a project three weeks ago. The client returns with a follow-up. What enables the agent to maintain context?

A A larger context window in the LLM.

B Long-term memory via vector stores or knowledge graphs.

C Saving the chat history in a text file.

D Re-running the original prompt.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A (larger context window) is limited—context windows don't persist across sessions. Three weeks of data won't fit.

C (text file) is naive. Raw text isn't efficiently retrievable for semantic context.

D (re-running prompts) doesn't provide historical context—just recreates the initial state.

WHY B IS CORRECT:

The document explicitly states "Remember: Long-term memory (via vector stores or knowledge graphs) ensures the agent learns from experience and maintains context over weeks or months."

19 AGENTIC TRANSFORMATION - HOW WE TRANSACT

SCENARIO

The document describes a future transaction model. What is it?

A Human buyers negotiating with AI customer service.

B Agent-to-Agent commerce where buying agents negotiate with selling agents.

C Humans negotiating through AI translation services.

D Blockchain-based smart contracts replacing all human negotiation.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A keeps humans in the loop for every transaction—not the described transformation.

C uses AI as a translation layer, not as an autonomous negotiating party.

D (blockchain) isn't mentioned in the source material.

WHY B IS CORRECT:

The document states: "How We Transact: We will see 'Agent-to-Agent' commerce, where your personal buying agent negotiates with a company's selling agent." This is autonomous commerce—agents transacting with each other.

20 WHAT PREDICTIVE AI COULD NOT DO

SCENARIO

A data scientist brags about their predictive model: "It forecasts churn with 95% accuracy!" Why does the Agent Factory thesis consider this insufficient?

A 95% accuracy is too low for production systems.

B Predictive AI tells you what might happen but cannot do anything about it—it's passive.

C Predictive AI only works with structured data, not text.

D Predictive AI requires too much training data.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A (accuracy threshold) misses the fundamental limitation.

C (data type) is an implementation detail, not a conceptual limitation.

D (training data) is a practical concern, not a philosophical one.

WHY B IS CORRECT:

The document explicitly states about Predictive AI: "it was passive—it could tell you what might happen, but it couldn't do anything about it." Predicting churn is valuable, but an agent would predict AND automatically execute a retention campaign.

SECTION C

GENERAL vs. CUSTOM AGENTS

> *"The strategist knows which tool to use. The fool uses the same hammer for every nail."*

21 THE CONSULTANT ANALOGY

SCENARIO

You need to debug a production system crash with an unknown root cause. Which agent type, and why?

A Custom Agent—it's more reliable for production systems.

B General Agent—like hiring a consultant, you tell them the problem and they figure out the path to solve it.

C Custom Agent—it can be pre-configured for crash debugging.

D Neither—humans should handle production emergencies.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A assumes reliability matters most here. But the root cause is UNKNOWN. Custom agents need defined workflows—they can't explore an undefined problem space.

C assumes you can pre-configure for unknown problems. This is contradictory. If the problem is unknown, you can't have a pre-built workflow for it.

D ignores the documented capability of General Agents for exactly this purpose.

WHY B IS CORRECT:

The document explicitly states General Agents are for "Complex debugging where the root cause is unknown" and uses the analogy: "Imagine hiring a high-priced Management Consultant or a Senior Engineer. You don't tell them HOW to fix the problem; you just tell them WHAT the problem is."

22 THE ASSEMBLY LINE ANALOGY

SCENARIO

You need to classify 5,000 support tickets into 10 categories. Which agent type?

A General Agent—it can understand the nuance in each ticket.

B Custom Agent—like an assembly line machine, it does one thing perfectly at high volume.

C General Agent—it's more intelligent and will make fewer mistakes.

D Custom Agent—because General Agents can't process more than 100 items.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A overvalues nuance for a standardized classification task. The categories are defined; this is repetitive, not novel.

C confuses intelligence with appropriateness. A General Agent's intelligence is wasted (and expensive) on a standardized, repetitive task.

D invents a false limitation about General Agents.

WHY B IS CORRECT:

The document states Custom Agents are for "High-volume tasks (e.g., classifying 5,000 support tickets)" and uses the analogy: "This is like building a specialized machine on a factory assembly line. It does one thing—like tightening a bolt or processing an invoice—perfectly, thousands of times a day."

23 THE DECISION MATRIX - TASK TYPE

SCENARIO

Task A: "Figure out why Q3 revenue dropped." Task B: "Process these 1,000 invoices the same way." Match to agent types.

A Task A → Custom, Task B → General

B Task A → General, Task B → General

C Task A → General, Task B → Custom

D Task A → Custom, Task B → Custom

CORRECT ANSWER **C**

WHY THE OTHERS FAIL:

A reverses the logic. "Figure out why" is novel problem-solving (General). "Process the same way" is standardized repetition (Custom).

B uses General for everything which ignores cost optimization for repetitive tasks.

D uses Custom for everything which would fail on the novel investigation task.

WHY C IS CORRECT:

The decision matrix states: General Agents handle "Novel, Problem-Solving: 'Figure out why this happened.'" Custom Agents handle "Repetitive, Standardized: 'Do this 100 times exactly the same way.'"

24 END USER CONSIDERATION

SCENARIO

You're building an agent for non-technical customers who need predictable, simple interactions. Which type?

A General Agent—its flexibility can adapt to any user.

B Custom Agent—it provides simple UI and predictable results for non-technical users.

C General Agent—customers prefer CLI interfaces.

D Custom Agent—because non-technical users can't afford General Agents.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A's "flexibility" is actually a problem for non-technical users who need predictability, not open-ended capability.

C is exactly backwards. The document says General Agents suit "Developers / Technical Staff: Comfortable with CLI." Non-technical users are NOT comfortable with CLI.

D (affordability) isn't the documented reason—user experience is.

WHY B IS CORRECT:

The decision matrix states Custom Agents suit "Non-Technical / Customers: Need a simple UI and predictable results."

25 ERROR TOLERANCE

SCENARIO

You're deploying an agent that will run unsupervised at 3 AM processing financial data. Which agent type and why?

A General Agent—it can self-correct if errors occur.

B Custom Agent—it has low error tolerance design, meaning it's built for reliable unsupervised operation.

C General Agent—financial data requires high intelligence.

D Custom Agent—because it's cheaper to run at night.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A misunderstands error tolerance. "Self-correction" is a General Agent feature, but the ASSUMPTION is a human is reviewing. Unsupervised operation needs LOW error tolerance (high reliability).

C conflates intelligence with reliability. Intelligence doesn't mean it won't make errors when unsupervised.

D (cost timing) is irrelevant to agent type selection.

WHY B IS CORRECT:

The decision matrix shows: General Agents have "High error tolerance: A human is usually reviewing the output." Custom Agents have "Low error tolerance: The system must run reliably without supervision." For unsupervised 3 AM operation, you need the low error tolerance design.

26 COST SENSITIVITY

SCENARIO

Your client says: "Every penny matters. We're processing 10,000 tasks per day and can't afford expensive per-task costs." Which agent type?

A General Agent—it's more efficient per task due to higher intelligence.

B Custom Agent—it's designed for volume optimization where cost per task matters.

C General Agent—the value of solving problems justifies any cost.

D Neither—10,000 tasks per day is too many for any agent.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A inverts the cost logic. General Agents are "computationally expensive per task"—the opposite of what this client needs.

C ignores the client's explicit constraint. They said cost matters—this answer dismisses their requirement.

D invents a volume limitation that doesn't exist.

WHY B IS CORRECT:

The decision matrix states: General Agents suit "Low cost sensitivity: The value of solving the problem is high; token cost is secondary." Custom Agents suit "High cost sensitivity: You need volume optimization; cost per task matters."

27 IMPLEMENTATION TIME

SCENARIO

Your startup needs an agent working by Friday. It's Tuesday. Which agent type is feasible?

A Custom Agent—you can design, code, test, and deploy in 3 days.

B General Agent—you can install, authenticate, and run immediately.

C Custom Agent—the SDK provides templates that take 2 hours to configure.

D General Agent—but only if you already have the infrastructure.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A contradicts the documented timeline. Custom Agents require "Weeks: Requires design, coding, testing, and deployment." Not 3 days.

C invents a claim about 2-hour SDK templates not in the source material.

D adds a false constraint. General Agents are described as instant—no special infrastructure mentioned.

WHY B IS CORRECT:

The decision matrix states: General Agents have "Instant implementation: Install, authenticate, and run." Custom Agents require "Weeks: Requires design, coding, testing, and deployment."

28 THE OODA LOOP ATTRIBUTION

SCENARIO

Which agent type explicitly operates using the OODA Loop (Observe, Orient, Decide, Act)?

A Custom Agents—they need OODA for reliability.

B General Agents—they use OODA for complex problem-solving.

C Both equally use OODA.

D Neither—OODA is a military concept, not an AI concept.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A misattributes OODA. Custom Agents follow defined workflows, not open-ended OODA loops.

C dilutes the distinction. The document specifically associates OODA with General Agents.

D ignores the explicit reference in the source material.

WHY B IS CORRECT:

The document states General Agents operate "in a tight OODA Loop (Observe, Orient, Decide, Act). It is computationally expensive per task but invaluable for non-routine work where the path to the solution is unclear."

29 CUSTOMER-FACING INTERACTIONS

SCENARIO

You need an agent for customer service that must follow strict scripts and never deviate. Which type?

A General Agent—it can understand customer nuance better.

B Custom Agent—it's designed for strict adherence to scripts and safety.

C General Agent—customers prefer intelligent responses.

D Custom Agent—because it's cheaper.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A values flexibility over the stated requirement (strict scripts, never deviate).

C assumes customer preference without evidence—and ignores the explicit "strict adherence" requirement.

D (cost) isn't the primary reason here—safety and script adherence are.

WHY B IS CORRECT:

The document states Custom Agents are best for "Customer-facing interactions where safety and strict adherence to scripts are required."

30 AD-HOC ANALYSIS

SCENARIO

The CEO walks in and asks: "Why did churn spike in Q3? I need an answer by lunch." Which agent type handles this?

A Custom Agent—create a churn analysis workflow quickly.

B General Agent—it handles ad-hoc data analysis with unclear paths to solution.

C Custom Agent—because it can query the database faster.

D Neither—this requires a human data analyst.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A requires time you don't have. Custom Agents take weeks to build. The CEO needs answers by lunch.

C confuses query speed with problem-solving capability. The challenge isn't speed—it's figuring out WHAT to query.

D ignores the documented capability of General Agents for exactly this scenario.

WHY B IS CORRECT:

The document explicitly lists General Agents as best for "Ad-hoc data analysis (e.g., 'Why did churn spike in Q3?')." This is a novel question requiring exploration, not a defined workflow.

SECTION D

CODE AS INTERFACE & MCP

> *"A conductor doesn't play every instrument. They orchestrate the symphony."*

31 THE CLAUDE CODE MISCONCEPTION

SCENARIO

A developer says: "Claude Code is just a coding assistant like Cursor—it helps me write Python faster." Why is this view "dangerously narrow" according to the document?

A Claude Code uses TypeScript, not Python.

B Claude Code is an Autonomous Problem Solver that uses code as its tool, not a role-bound pair programmer.

C Claude Code is slower than Cursor.

D Claude Code can't actually write code—it only explains code.

CORRECT ANSWER B

WHY THE OTHERS FAIL:

A (language specificity) isn't the conceptual distinction being made.

C (speed comparison) is irrelevant to the architectural difference.

D is factually wrong—Claude Code clearly writes code.

WHY B IS CORRECT:

The document explicitly states this misconception is "dangerously narrow" and clarifies: "Coding Agents (e.g., Cursor): These act as 'pair programmers.' They live in your IDE and help you complete specific coding tasks. They are role-bound. General Agents (Claude Code): These operate at the system level. They can solve problems across any domain—finance, marketing, ops—by writing code to interface with those domains."

32 CODE AS UNIVERSAL INTERFACE - THE CEO EXAMPLE

SCENARIO

A CEO asks: "Why did sales drop in Q3?" A standard chatbot hallucinates an answer. What does a General Agent do differently?

It asks the CEO clarifying questions until it can guess better.

A

B It writes SQL to fetch real data, Python to visualize trends, analyzes the chart, and delivers an evidence-based answer.

C It says "I don't have access" and asks for database credentials.

D It summarizes what it learned during training about typical sales drop causes.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A is just better guessing—still no real data.

C gives up instead of using tools to solve the problem.

D (training knowledge) is exactly what leads to hallucination—not grounded in this company's actual data.

WHY B IS CORRECT:

The document describes exactly this: "A General Agent: 1) Writes a SQL query to fetch real sales data. 2) Writes a Python script (using matplotlib/pandas) to visualize the trend. 3) Analyzes the chart visually. 4) Delivers the Answer: 'Sales dropped because of 40% churn in the Enterprise sector, specifically in EMEA.'"

33 THE COGNITIVE LEAP

SCENARIO

An early coding agent made a mistake. It couldn't recover. A General Agent makes the same mistake. What happens?

A Both fail—AI can't recover from mistakes.

B The General Agent uses OODA to observe the error, decide on correction, act, and fix itself.

C The General Agent asks a human to fix it.

D The General Agent restarts from scratch.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A ignores the documented capability of self-correction.

C requires human intervention—which defeats autonomy.

D (restart) is brute force, not intelligent recovery.

WHY B IS CORRECT:

The document explicitly contrasts them: "Early Coding Agents (Predictive): They were great at autocomplete but terrible at recovery. If they made a mistake, they couldn't fix it. General Agents (Reasoning Loop): Observe: 'I see a Connection Refused error.' Decide: 'Check container status.' Act: 'docker ps.' Correct: 'Container stopped. Restart it.' This self-correction capability is the breakthrough."

34 MCP AS UNIVERSAL CONNECTOR

SCENARIO

You want your General Agent to manage Slack channels AND audit Salesforce pipelines AND reconcile QuickBooks. What enables this?

- **A** You need to build three separate agents, one for each tool.
- **B** MCP—you plug in MCP servers for each tool and the agent instantly gains those capabilities.
- **C** You need to fine-tune the LLM on each tool's documentation.
- **D** It's impossible—agents can only connect to one tool at a time.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A multiplies complexity unnecessarily when a single agent can be extended.

C (fine-tuning) is expensive and slow compared to MCP's plug-and-play model.

D invents a false limitation.

WHY B IS CORRECT:

The document states: "If Code is the interface, MCP is the universal connector. By plugging in MCP servers, you can expand a General Agent's role instantly: Communications Manager (Slack MCP), RevOps Specialist (Salesforce MCP), Financial Auditor (Xero/QuickBooks MCP)."

35 WHY CODE, NOT NATURAL LANGUAGE ALONE?

SCENARIO

Someone argues: "If agents understand English, why do they need to write code? Just have them explain what to do."

A They're right—code is unnecessary if English prompts are good enough.

B Wrong—code is the Universal Interface for reality interrogation. Explanations don't change the world; executed code does.

C Code is faster than English.

D English has too many ambiguities for AI to understand.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A misses the action/execution gap. Explaining what to do doesn't DO it.

C (speed) isn't the fundamental reason.

D (ambiguity) is a real concern but not the core argument here.

WHY B IS CORRECT:

The document calls code "the Universal Interface for reality interrogation." The point isn't understanding—it's ACTION. "The agent used code not to build an app, but to answer a business question" by actually querying real data, not just discussing what data might exist.

36 THE ROLE TRANSFORMATION VIA MCP

SCENARIO

You plug the Salesforce MCP + a Sales Skill into your agent. What does it become?

A A chatbot that can discuss Salesforce features.

B A RevOps Specialist that can audit pipelines and score leads.

C A Salesforce administrator that can configure the CRM.

D A Sales trainer that teaches humans how to use Salesforce.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A (discuss features) is passive chatbot behavior, not agentic action.

C (administrator) conflates CRM configuration with sales operations analysis.

D (trainer) is a different function entirely.

WHY B IS CORRECT:

The document explicitly states: "RevOps Specialist: Plug in the Salesforce MCP + a Sales Skill. The agent can now audit pipelines and score leads."

37 CODING AGENT VS. GENERAL AGENT - IDE SCOPE

SCENARIO

Where does Cursor live and operate? Where does Claude Code live and operate?

A Both live in the IDE.

B Cursor lives in the IDE (role-bound); Claude Code operates at system level (cross-domain).

C Cursor operates at system level; Claude Code lives in the IDE.

D Neither requires an IDE—both are cloud-only.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A misses the crucial scope difference.

C reverses the documented relationship.

**D contradicts the explicit IDE reference for Cursor.

WHY B IS CORRECT:

The document states: "Coding Agents (e.g., Cursor): These act as 'pair programmers.' They live in your IDE and help you complete specific coding tasks. They are role-bound. General Agents (Claude Code): These operate at the system level."

38 PREDICTIVE VS. REASONING AGENTS

SCENARIO

An early coding agent worked by: "Based on the last 10 lines of code, what is the most likely next line?" This is which type of intelligence?

A Reasoning-based intelligence.

B Probability-based prediction without recovery capability.

C Agentic intelligence with OODA loops.

D Generative creativity.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A (reasoning) implies analysis and correction—which these models lacked.

C (OODA) is the newer paradigm, not the old one being described.

D (creativity) isn't the focus—it's autocomplete, not creation.

WHY B IS CORRECT:

The document describes early coding agents as "Predictive: These models worked on probability: 'Based on the last 10 lines of code, what is the most likely next line?' They were great at autocomplete but terrible at recovery."

39 THE TROJAN HORSE CONCEPT

SCENARIO

Why does the document call Claude Code a "Trojan Horse"?

A It secretly installs malware.

B It appears to be a coding tool but is actually an Autonomous Problem Solver that uses code to solve ANY domain problem.

C It was created by a Greek company.

D It requires hidden authentication to use.

CORRECT ANSWER B

WHY THE OTHERS FAIL:

A (malware) is absurd.

C (Greek) is irrelevant wordplay.

D (authentication) isn't the metaphor's meaning.

WHY B IS CORRECT:

The section is titled "The 'Trojan Horse'" and explains that the "widespread misconception that tools like 'Claude Code' are merely 'coding assistants'" is wrong—it's actually an "Autonomous Problem Solver that happens to use code as its primary tool." The Trojan Horse metaphor means it enters as one thing (coding tool) but is actually something much more powerful (autonomous problem solver).

40 WHAT MCP ACTUALLY ENABLES

SCENARIO

Without MCP, what would an agent need to do to connect to a new tool?

A Nothing—agents connect to tools automatically.

B Custom integration code for each tool, preventing easy extensibility.

C Just ask the LLM to figure it out.

D Download the tool's mobile app.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A ignores the integration problem MCP solves.

C (LLM figuring it out) doesn't provide actual connectivity—just guessing.

D (mobile app) is irrelevant.

WHY B IS CORRECT:

The document describes MCP as the "universal connector"—implying that without a universal standard, you'd need custom integrations for each tool. MCP provides "infinite extensibility" by standardizing how agents plug into external systems.

SECTION E

THE AGENT FACTORY WORKFLOW

> *"A factory without a blueprint produces chaos, not products."*

41 THE FACTORY CONCEPT

SCENARIO

What IS the "Agent Factory" conceptually?

A A company that manufactures robots.

B A methodology where General Agents (builders) manufacture Custom Agents (results) using Specs.

C A cloud platform for hosting agents.

D A training program for AI engineers.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A confuses digital agents with physical robots.

C (cloud platform) is infrastructure, not the methodology itself.

D (training program) is adjacent but not the core concept.

WHY B IS CORRECT:

The document states: "The 'Agent Factory' is a methodology where we use General Agents (the Builder) to manufacture Custom Agents (the Result). We don't write the code for the custom agents manually; we orchestrate the builder to do it."

42 THE FOUR-STAGE WORKFLOW

SCENARIO

Put these in the correct order: Manufacturing, Result, Spec, Builder.

A Builder → Spec → Manufacturing → Result

B Spec → Builder → Manufacturing → Result

C Result → Manufacturing → Builder → Spec

D Manufacturing → Spec → Result → Builder

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A starts with Builder but you can't build without knowing what to build (Spec first).

C starts with Result which is the output, not the input.

D has no logical flow.

WHY B IS CORRECT:

The document explicitly states: "1. The Spec (The Blueprint) → 2. The Builder (Claude Code) → 3. The Manufacturing (The Build Process) → 4. The Result (The Product)."

43 THE SPEC - WHAT IT CONTAINS

SCENARIO

A good Spec for an agent that automates Q3 Financial Audits would include:

A Just the goal: "Automate Q3 audits."

B Goal + specific actions: "Read PDFs from this folder, extract tax IDs, verify against government API, log discrepancies to CSV."

C Only the code to be generated.

D Only the budget and timeline.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A is too vague. "Automate audits" gives the builder no specific requirements.

C (code only) defeats the purpose—you write specs so AI generates the code.

D (budget/timeline) are project management details, not technical specs.

WHY B IS CORRECT:

The document gives this exact example: "I need an agent that automates Q3 Financial Audits. It must read PDFs from this folder, extract tax IDs, verify them against the government API, and log discrepancies to a CSV."

44 WHAT CLAUDE CODE DOES AS BUILDER

SCENARIO

When Claude Code receives a Spec, what does it do?

- **A** It sends the Spec to Anthropic for manual implementation.
- **B** It reads requirements, scans API documentation, identifies tools, generates folder structure, writes SKILL.md and scripts, and iteratively tests.
- **C** It just summarizes the Spec in simpler terms.
- **D** It asks for more budget before starting.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A (manual implementation) defeats the purpose of automation.

C (summarization) adds no value—it doesn't build anything.

D (budget requests) is absurd.

WHY B IS CORRECT:

The document states: "Claude Code acts as the engineer. It generates the folder structure, writes the SKILL.md (the instructions), and writes the supporting Python scripts to handle the PDF parsing and API calls. It iteratively tests the code until it works."

45 THE RESULT OF THE FACTORY

SCENARIO

After the Agent Factory process completes, what do you have?

A A prototype that needs human developers to finish.

B A production-ready Custom Agent or Custom Skill deployable for 24/7 autonomous operation.

C Documentation explaining how to build the agent manually.

D A request for more training data.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A (prototype needing humans) contradicts the "production-ready" promise.

C (documentation) is useful but not the primary output.

D (training data request) doesn't fit the workflow.

WHY B IS CORRECT:

The document states: "A production-ready Custom Agent or Custom Skill is born. You can now deploy this asset to run 24/7 without your involvement."

46 SPEC-DRIVEN DEVELOPMENT (SDD)

SCENARIO

What is Spec-Driven Development fundamentally opposing?

A Test-Driven Development

B Vibe Coding—vague prompts leading to drift and unmaintainable software

C Object-Oriented Programming

D Agile methodology

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A (TDD) is actually compatible with SDD, not opposed.

C (OOP) is a coding paradigm, not a process methodology.

D (Agile) is a project management framework, not what SDD opposes.

WHY B IS CORRECT:

The document explicitly states: "Many developers today practice 'Vibe Coding'—they vaguely prompt the AI, get a mediocre result, re-prompt, and iterate endlessly. This leads to 'drift,' where the final code doesn't match the original requirements, and creates unmaintainable software. Spec-Driven Development (SDD) reverses this."

47 THE SDD WORKFLOW

SCENARIO

What's the correct SDD workflow?

A AI Executes → Spec → Quality Results

B Quality Results → AI Executes → Spec

C Spec First → AI Executes → Quality Results

D Spec → Quality Results → AI Executes

CORRECT ANSWER **C**

WHY THE OTHERS FAIL:

A starts with AI executing without knowing what to execute.

B starts with results which are outputs, not inputs.

D puts results before execution which is impossible.

WHY C IS CORRECT:

The document explicitly states: "The Workflow: Spec First → AI Executes → Quality Results."

48 THE BLUEPRINT FOR A PERFECT SPEC

SCENARIO

A Senior-level Spec must include which elements?

- **A** Just the goal and deadline.
- **B** Identity, Context, Logic, Success Trigger, Output Standard, Error Protocol.
- **C** Only the database schema.
- **D** Only the expected revenue.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A is incomplete—missing technical requirements.

C (schema only) is one detail, not a complete spec.

D (revenue) is a business metric, not a technical spec element.

WHY B IS CORRECT:

The document lists: "Identity (Persona + Tone), Context (MCP Tool Access + Knowledge Base), Logic (Workflow + Guardrails), Success Trigger, Output Standard (JSON schema or template), Error Protocol (fallback actions)."

49 THE "ENGINE" OF THE FACTORY

SCENARIO

What serves as the "Engine" of the Agent Factory?

A Custom Agents—they do the actual work.

B General Agents (like Claude Code)—they transform specs into code assets.

C The database that stores all the data.

D The end users who provide requirements.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A (Custom Agents) are the OUTPUT, not the engine.

C (database) is storage, not the transformation engine.

D (end users) provide inputs but don't transform them.

WHY B IS CORRECT:

The document states: "General Agents (like Claude Code) serve as the 'Engine' of the factory. They utilize Spec-Driven Development (SDD) to transform human intent into executable code assets."

50 SPEC KIT PLUS COMPONENTS

SCENARIO

What does Spec Kit Plus include?

A Only SPEC.md

B SPEC.md, Vertical Sub-agent templates, Prompt History logs, TDD setups

C Only deployment scripts

D Only marketing materials

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A is incomplete.

C (deployment scripts) are infrastructure, not spec methodology.

D (marketing) is irrelevant to technical spec development.

WHY B IS CORRECT:

The document states Spec Kit Plus includes: "SPEC.md: The master document containing Feature Description, Technical Requirements, and Acceptance Criteria. Components: Includes templates for Vertical Sub-agents, Prompt History logs, and Test-Driven Development (TDD) setups."

SECTION F

AGENT SKILLS & PROGRESSIVE DISCLOSURE

> *"The master craftsman knows that the best work requires the right tools—loaded just in time."*

51 SKILLS DEFINITION

SCENARIO

What ARE Agent Skills fundamentally?

A The LLM's pre-trained abilities from its training data.

B Organized file collections packaging composable, procedural knowledge—like cartridges loaded into an agent.

C Certifications that developers earn.

D The agent's personality settings.

CORRECT ANSWER B

WHY THE OTHERS FAIL:

A (pre-trained abilities) confuses model capabilities with external skill modules.

C (certifications) conflates human credentials with agent capabilities.

D (personality) is one aspect of skills but not the full definition.

WHY B IS CORRECT:

The document states: "Skills are the fundamental unit of knowledge in the Agent Factory. They are organized collections of files that package composable, procedural knowledge. Think of them as 'cartridges' you load into an agent to give it a specific capability."

52 SKILL FOLDER STRUCTURE

SCENARIO

You're building a skill folder. What's the correct structure?

A Just a single README.txt file.

B SKILL.md (brain), docs.md (manual), rules/slide-decks.md (rules), script.py (hands)

C Only Python files.

D Only JSON configuration files.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A is too minimal and uses the wrong format.

C (Python only) misses the instruction files that guide when/how to use scripts.

D (JSON only) misses the markdown instructions and executable scripts.

WHY B IS CORRECT:

The document shows: "SKILL.md: The 'Brain.' Contains YAML frontmatter and instructions. docs.md: The 'Manual.' Detailed documentation. slide-decks.md: The 'Rules.' Specific domain rules. apply_template.py: The 'Hands.' Executable script that performs actual work."

53 WHAT SKILL.MD CONTAINS

SCENARIO

What does the SKILL.md file contain?

A Only the Python code.

B YAML frontmatter (metadata) and high-level instructions on when and how to use the skill.

C Only the API credentials.

D The full conversation history.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A (Python code) goes in separate script files.

C (credentials) should never be stored in SKILL.md.

D (conversation history) is runtime data, not skill definition.

WHY B IS CORRECT:

The document states: "SKILL.md: The 'Brain.' Contains the YAML frontmatter (metadata) and high-level instructions on when and how to use the skill."

54 WHY SCRIPTS IN SKILLS?

SCENARIO

Why do Skills include Python scripts instead of having the LLM figure out the logic?

- **A** LLMs can't run code.
- **B** Scripts execute deterministic code—we don't want the LLM guessing, we want perfect execution.
- **C** Scripts are faster to load.
- **D** Scripts cost less to store.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A is wrong—LLMs can generate and reason about code.

C (loading speed) isn't the primary reason.

D (storage cost) is negligible and not the point.

WHY B IS CORRECT:

The document states: "The power of a Skill lies in its ability to execute deterministic code. We don't want the LLM to guess how to modify a PowerPoint file; we want it to run a script that does it perfectly."

55 PROGRESSIVE DISCLOSURE STRATEGY

SCENARIO

What is Progressive Disclosure in the context of Agent Skills?

A Slowly revealing the agent's identity to users.

B A technique to manage context window limits by loading small index files at startup and detailed docs only when needed.

C Gradually giving agents more permissions over time.

D A user interface design pattern for mobile apps.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A (identity revelation) is unrelated.

C (permissions) conflates access control with loading strategy.

D (mobile UI) is a different domain.

WHY B IS CORRECT:

The document states: "We use a technique called Progressive Disclosure to manage context window limits and token costs. SKILL.md (~100 tokens) is always loaded at startup as an index. Detailed Docs are only read if the agent decides it needs them."

56 TOKEN CONSUMPTION - SCRIPTS

SCENARIO

How many tokens do Python scripts consume in the agent's context?

A About 1,000 tokens per file.

B Zero tokens—scripts are executed, not read by the LLM.

C It depends on the script length.

D About 100 tokens per function.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A, C, D all assume scripts are read into context. They are not.

WHY B IS CORRECT:

The document explicitly states: "Scripts (0 tokens always): Scripts like helper.py are executed by the Python interpreter. Their source code is never read by the LLM; only their output (success/fail) enters the context."

57 TOKEN EFFICIENCY COMPARISON

SCENARIO

Loading raw MCP tool definitions costs how many tokens vs. Skills + Scripts?

A MCP tools: ~100 tokens; Skills: ~14,000 tokens

B MCP tools: 14,000-80,000+ tokens (~41% of context); Skills: ~100 tokens (~3% of context)

C Both cost about the same.

D Skills cost more because they include documentation.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A reverses the numbers.

C (same cost) contradicts the documented 80-98% reduction.

D (Skills cost more) is exactly backwards.

WHY B IS CORRECT:

The document states: "MCP Tools Only: 14,000–80,000+ tokens just to start, ~41% of context window. Skills + Scripts: ~100 tokens, ~3% of context. This results in an 80-98% token reduction."

58 THE AGENT SKILLS OPEN STANDARD

SCENARIO

When was Agent Skills announced as an independent open standard, and who announced it?

A 2024, announced by OpenAI

B December 19, 2025, announced by agentskills.io

C 2023, announced by Google

D 2026, announced by Anthropic

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

All other options have wrong dates or organizations.

WHY B IS CORRECT:

The document explicitly states: "On December 19, 2025, agentskills.io announced Agent Skills as an independent open standard."

59 CLAUDE CODE SKILLS DIRECTORY FORMAT

SCENARIO

Where does Claude Code natively store skills?

A `/skills/SKILL.md`

B `.claude/skills/SKILL.md`

C `.openai/skills/`

D `/agent_skills/`

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A, C, D don't match the documented format.

WHY B IS CORRECT:

The document's adoption table shows Claude Code format as: `.claude/skills/SKILL.md`

60 WHAT IS AGENTS.MD?

SCENARIO

You see a file called AGENTS.md in a project root. What is it?

A A list of all human team members.

B The robots.txt for AI Agents—universal instructions about project structure, conventions, and documentation locations.

C A deployment configuration file.

D A database schema file.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A (team members) confuses agents with humans.

C (deployment) is a different file type.

D (schema) is unrelated.

WHY B IS CORRECT:

The document states: "AGENTS.md: This is the robots.txt for AI Agents. Purpose: A universal instruction file placed in the root of a project. Usage: It tells any visiting AI agent how the project is structured, coding conventions to follow, and where documentation lives."

SECTION G

DIGITAL FTE & MONETIZATION

> *"If you can't price it, you can't sell it. If you can't sell it, it's a hobby, not a business."*

61 DIGITAL FTE DEFINITION

SCENARIO

What IS a Digital FTE?

A A freelance human who works remotely full-time.

B The commercial packaging of an AI agent as the equivalent of a full-time human employee.

C A software license that costs the same as a human employee.

D A robot that works in a factory.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A (freelance human) is a human, not an AI agent.

C (software license) is a pricing model, not a definition.

D (robot) conflates digital agents with physical robots.

WHY B IS CORRECT:

The document states: "A Digital Full-Time Equivalent (FTE) is the commercial packaging of an AI agent. It represents the transition from selling software tools to selling labor."

62 HOURS PER WEEK COMPARISON

SCENARIO

A Human FTE works X hours/week. A Digital FTE works Y hours/week. What are X and Y?

A Human: 60, Digital: 80

B Human: 40, Digital: 168 (24/7)

C Human: 40, Digital: 40

D Human: 168, Digital: 40

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A has wrong numbers.

C assumes equal availability—missing the 24/7 advantage.

D reverses the values.

WHY B IS CORRECT:

The document's ROI table explicitly shows: "Human FTE: 40 hours/week. Digital FTE: 168 hours/week (24/7)."

63 COST PER TASK REDUCTION

SCENARIO

The document suggests highlighting which metric when pitching Digital FTEs?

A Total annual savings

B 85-90% cost-per-task reduction

C Number of agents deployed

D Lines of code written

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A (annual savings) is a derived metric, not the primary pitch.

C (agent count) doesn't translate to business value.

D (lines of code) is a vanity metric with no business meaning.

WHY B IS CORRECT:

The document states: "Sales Strategy: When pitching this, highlight the Cost per Task reduction (85-90%). This metric bypasses IT budget constraints and taps into the much larger Operations/HR budgets."

64 MONETIZATION MODEL - DIGITAL FTE SUBSCRIPTION

SCENARIO

The Digital FTE subscription model charges $1,000/month. What value does this provide the client?

- **A** Access to source code.
- **B** Hands-off automation—client treats it like an employee salary.
- **C** Unlimited API calls.
- **D** Priority customer support.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A (source code) is a licensing concern, not the subscription value prop.

C (API calls) is a technical detail, not the business framing.

D (support) is a feature, not the core value proposition.

WHY B IS CORRECT:

The document states: "Digital FTE (Subscription): Model: Monthly fee for a fully managed agent. Value: Hands-off automation. The client treats it like an employee salary."

65 SUCCESS FEE MODEL

SCENARIO

A client only wants to pay when the agent delivers results. Which model?

A Digital FTE Subscription

B Success Fee / Outcome-Based

C White-Label License

D Developer License

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A (subscription) requires payment regardless of results.

C (white-label) is about IP ownership, not outcomes.

D (developer license) is for sub-modules, not outcome-based.

WHY B IS CORRECT:

The document states: "Success Fee (Outcome-Based): Model: Commission on results (e.g., $5 per qualified lead, or 2% of recovered tax savings). Value: High trust alignment. The client only pays when they win."

66 COCOUNSEL CASE STUDY - ACQUISITION VALUE

SCENARIO

CoCounsel was acquired for how much, and by whom?

A $100 Million by Microsoft

B $650 Million in cash by Thomson Reuters

C $1 Billion by Google

D $50 Million by OpenAI

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

All other options have wrong amounts or acquirers.

WHY B IS CORRECT:

The document states: "The company was acquired for $650 Million in cash by Thomson Reuters."

67 WHY COCOUNSEL PRICED AT $500/MONTH VS. $20/MONTH

SCENARIO

CoCounsel charged $500/month while basic legal tools charge $20/month. Why the 25x premium?

A CoCounsel used more expensive AI models.

B CoCounsel sold a "seat" (Digital Employee model) not a "tool"—it performed substantive legal work.

C CoCounsel had better branding.

D CoCounsel included customer support.

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A (model cost) doesn't justify 25x premium to customers.

C (branding) doesn't explain the value difference.

D (support) doesn't command 25x pricing.

WHY B IS CORRECT:

The document states: "Instead of selling it as a tool, they sold it as a 'seat' for $500/month (compared to $20/month for basic legal tools)... proving the immense value of the 'Digital Employee' model over the 'SaaS Tool' model."

SECTION H

SECURITY, ARCHITECTURE & WHEN NOT TO USE

> *"A master knows their tools' limits as well as their capabilities."*

68 WHEN NOT TO USE AI AGENTS

SCENARIO

Your CEO wants to deploy an agent for medical diagnoses. Based on the document, your response is:

A "Great idea—agents are perfect for healthcare."

B "No—irreversible high-stakes decisions like medical diagnoses are explicitly listed as inappropriate for agents."

C "Only if we use a Custom Agent."

D "Only if we add more guardrails."

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A ignores the explicit warning in the document.

C (Custom Agent) doesn't change the fundamental inappropriateness.

D (guardrails) can't fully mitigate irreversible high-stakes decisions.

WHY B IS CORRECT:

The document explicitly states: "When NOT to Use AI Agents: Irreversible High-Stakes Decisions: Medical diagnoses, legal judgments, or deploying code to production without review."

69 SECURITY FRAMEWORK - ENCRYPTION STANDARDS

SCENARIO

What encryption standards does the document specify for enterprise agent deployment?

A AES-128 at rest, SSL for transit

B AES-256 for data at rest, TLS 1.3 for data in transit, 90-day key rotation

C No encryption required

D Only TLS 1.2

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A has weaker standards (AES-128, SSL instead of TLS 1.3).

C (no encryption) would be a security disaster.

D (only TLS 1.2) misses at-rest encryption and is an older TLS version.

WHY B IS CORRECT:

The document states: "Data Encryption: AES-256 for data at rest, TLS 1.3 for data in transit, with key rotation every 90 days."

70 THE DUAL-BACKEND ARCHITECTURE

SCENARIO

The Agent Factory technical architecture uses a dual-backend system. What are the two backends?

A Frontend and Backend

B Deterministic Backend (Zero LLM for trivial tasks) and Hybrid Backend (LLM for reasoning tasks)

C SQL Backend and NoSQL Backend

D Cloud Backend and Edge Backend

CORRECT ANSWER **B**

WHY THE OTHERS FAIL:

A (frontend/backend) is a different architectural distinction.

C (SQL/NoSQL) is about database types, not the reasoning/deterministic split.

D (cloud/edge) is about deployment location, not processing type.

WHY B IS CORRECT:

The document states: "The Engine Room: A Dual-Backend System. Deterministic Backend (Zero LLM): Uses standard code for tasks that don't require reasoning (fetching data, formatting, math). Predictable, fast, cheap. Hybrid Backend: Calls the LLM only when reasoning is required. This optimizes cost by not wasting tokens on trivial tasks."

ANSWER KEY

Q1 **B**	Q2 **B**	Q3 **B**	Q4 **C**	Q5 **B**	Q6 **C**	Q7 **C**
Q8 **C**	Q9 **C**	Q10 **B**	Q11 **A**	Q12 **B**	Q13 **B**	Q14 **B**
Q15 **B**	Q16 **B**	Q17 **B**	Q18 **B**	Q19 **B**	Q20 **B**	Q21 **B**
Q22 **B**	Q23 **C**	Q24 **B**	Q25 **B**	Q26 **B**	Q27 **B**	Q28 **B**
Q29 **B**	Q30 **B**	Q31 **B**	Q32 **B**	Q33 **B**	Q34 **B**	Q35 **B**
Q36 **B**	Q37 **B**	Q38 **B**	Q39 **B**	Q40 **B**	Q41 **B**	Q42 **B**
Q43 **B**	Q44 **B**	Q45 **B**	Q46 **B**	Q47 **C**	Q48 **B**	Q49 **B**
Q50 **B**	Q51 **B**	Q52 **B**	Q53 **B**	Q54 **B**	Q55 **B**	Q56 **B**
Q57 **B**	Q58 **B**	Q59 **B**	Q60 **B**	Q61 **B**	Q62 **B**	Q63 **B**
Q64 **B**	Q65 **B**	Q66 **B**	Q67 **B**	Q68 **B**	Q69 **B**	Q70 **B**

SCORING GUIDE

Score Range	Level	Assessment
63-70 (90-100%)	MASTER LEVEL	You understand the architecture at first principles. Ready for the real exam.
56-62 (80-89%)	PROFICIENT	Strong understanding but some conceptual gaps. Review missed sections.
49-55 (70-79%)	DEVELOPING	Surface knowledge but lack deep understanding. Focus on WHY, not WHAT.
42-48 (60-69%)	BEGINNER	You're memorizing, not understanding. Start over with foundations.
Below 42 (<60%)	NOT READY	Intensive study needed. Don't take real exam until 90%+ here.

Final Note from Your PhD Examiner

If you scored below 90%, you fell for the traps. The traps exist because most people learn WHAT without learning WHY.

The "vibe coder" picks A because it sounds technical.
The memorizer picks C because they saw that word somewhere.
The **strategist** picks B because they understand the underlying architecture.

Be the strategist. The real exam will be harder.

Exam Created: January 2026 | Source: FTE Slides Content | Author: Zia Khan
Exam Designer: Afnan Zahed, PIAIC, Islamabad | LinkedIn | WhatsApp Group