ChatGPT

Selling Shovels in the AI Gold Rush: A Strategic, VC-Style Map Across Industries and Finance

Gold Rush Framework for the AI Era

What represents the "gold" in AI?
In prior rushes the "gold" was a scarce, high-value resource (gold nuggets; eyeballs/traffic on the internet; cloud elasticity and OPEX substitution in SaaS). In the AI era, the "gold" is **capturable economic surplus from automating and augmenting cognitive work**—measured as productivity gains, margin expansion, and new product revenue powered by models and agents. A widely cited estimate pegs generative AI's annual economic potential at **$2.6T–$4.4T** across dozens of use cases, which is a useful proxy for the scale of the prize. [1]

Who are the "miners"?
Miners are the actors taking **direct, high-variance bets** on extracting that surplus:

- **Frontier model builders** (training and serving foundation models), where training compute for notable models is rising rapidly and industry has become the dominant producer of notable models. [2]
- **Application-layer companies** embedding AI into products and workflows (horizontal copilots and vertical apps), competing on distribution, UX, and domain fit—often in crowded markets where differentiation can be thin and churn high because model inference is becoming cheaper and more accessible. [3]

Who are the "shovel sellers"?
Shovel sellers are businesses that profit **regardless of which miner wins**, by supplying inputs every miner needs:

- **Compute inputs** (accelerators, memory, packaging capacity, and the physical data-center buildout). Bottlenecks like advanced packaging and high-bandwidth memory are explicitly described as capacity constraints and "sold out/allocated" dynamics. [4]
- **Energy and grid capacity** enabling data centers; electricity demand from data centers is already material and projected to rise substantially. [5]
- **Cloud distribution** for compute (where the largest providers aggregate demand and supply). Enterprise cloud infrastructure spending is concentrated among the top providers, reinforcing "land ownership" dynamics. [6]
- **Trust, security, and governance tooling** that reduces deployment risk, aligned with emerging standards and vulnerability taxonomies. [7]

Who "owns the land/platforms"?
In classic rushes, landowners controlled scarce access (claims, rail hubs, ports). In AI, "land" is a bundle of scarce, chokepoint assets:

1. **Compute chokepoints**: advanced packaging capacity expansion plans (eg, CoWoS) and HBM supply dynamics make clear that capacity—not just chip design—is limiting output. [8]
2. **Power + interconnect**: data centers already consume hundreds of TWh globally and are projected to exceed 1,000 TWh in the next decade under baseline scenarios; this puts grids, permits, and energy procurement at the center of platform power. [5]
3. **Cloud distribution and metering**: the top cloud providers together capture a large majority of enterprise cloud infrastructure services spend, giving them pricing, product, and ecosystem control. [6]
4. **Regulatory perimeter**: rules define what "safe and compliant" AI deployment means, shifting spend toward compliance-grade tooling and auditability (for example, the EU AI Act's staged applicability). [9]

Historical parallel (why "shovels" win):
During the California Gold Rush, merchants and service providers famously built durable fortunes by selling essentials and financial services rather than prospecting. Authoritative sources explicitly cite entity["company","Levi Strauss & Co.","denim maker, san francisco ca"] and entity["company","Wells Fargo & Co.","bank, san francisco ca"] as examples of gold-rush-era "support" winners who served miners' needs. [10]
The internet and cloud booms repeated the pattern: when a platform layer aggregates demand, durable profits tend to accrue to infrastructure, tooling, and distribution owners. The present concentration of cloud market share among top providers mirrors that "platform-as-land" structure. [6]

The AI Shovel Seller Landscape Across Industries

The most investable "shovel" categories share three traits: (1) **non-discretionary demand** as AI adoption expands, (2) **multi-tenant relevance** (many miners must buy), and (3) either **structural scarcity** (hard to supply) or **structural switching costs** (hard to replace).

Cross-industry shovel categories (with VC-grade cut)

Below is a compact map emphasizing *why the category is a shovel*, *what drives demand*, *where margins and moats come from*, and *what is still open*.

Shovel category	Why it's a "shovel" (wins regardless of model/app winner)	Demand / market proof points	Profitability potential	Entry barriers	Leading incumbency signals	Underserved opportunities still open
Compute supply chain: accelerators, memory, packaging	Every frontier/ enterprise AI deployment consumes accelerators + memory + advanced packaging; shortages shift bargaining power upstream	Advanced packaging is repeatedly described as capacity-limiting; HBM supply has been described as largely allocated/sold out in prior cycles; market-share and next-gen ramp dynamics show persistent scarcity [4]	High when scarce; often oligopolistic economics at the tightest chokepoints	Extreme capex, process know-how, geopolitics	NVIDIA [11] posts enormous data-center revenue and high gross margins, illustrating concentrated profit pools when supply is constrained [12]	"Software-defined capacity" around the bottlenecks: packaging allocation optimization, yield analytics, secure supply-chain provenance for chips/memory
Cloud + managed AI compute	Aggregates enterprise demand; meters usage; becomes default procurement channel	Top providers jointly hold a dominant share of cloud infra services spend; market scale cited at ~$107B in a quarter [6]	Attractive at scale via utilization + services attach	Massive capex + global sales + trust	Concentration among top providers (market-share data) [6]	Neutral "multi-cloud AI control plane": portability, GPU scheduling, inference caching, FinOps for LLM/agent workloads

Shovel category	Why it's a "shovel" (wins regardless of model/app winner)	Demand / market proof points	Profitability potential	Entry barriers	Leading incumbency signals	Underserved opportunities still open
Data-center physical infrastructure + energy	AI is ultimately "electricity turned into tokens"; growth pushes spend into power, cooling, siting	Data centers consume ~415 TWh (~1.5% of global electricity) and demand is projected to rise materially over the decade [5]	Strong in constrained geographies; recurring services possible	Permits, supply chain, utility interconnects	Energy constraints are now first-order in data-center planning (IEA projections) [13]	Grid-interconnect acceleration tooling; workload-flexibility markets; "AI demand response" and carbon-aware routing products
LLM/agent security and safety tooling	As AI penetrates workflows, attackers gain new vectors (prompt injection, data leakage, model DoS); defenders must tool up	OWASP explicitly catalogs LLM risks (prompt injection, training data poisoning, model DoS, supply chain vulnerabilities) for LLM applications [14]	High if embedded in enterprise security budgets	Need deep security + ML expertise; trust barrier	Security risk taxonomies now exist (OWASP), enabling category creation [15]	Agent permissioning systems; runtime control for tool-use; red teaming-as-a-service for regulated industries

Shovel category	Why it's a "shovel" (wins regardless of model/app winner)	Demand / market proof points	Profitability potential	Entry barriers	Leading incumbency signals	Underserved opportunities still open
AI governance, risk, and audit infrastructure	Regulators and boards demand lifecycle risk management; this forces spend into control frameworks and evidence	National Institute of Standards and Technology [16] publishes AI RMF 1.0 as a widely referenced risk management framework [17]; EU AI Act timeline makes compliance deadlines concrete [9]	"Compliance-grade" products can be sticky and high-ACV	Requires legal/reg + engineering; procurement cycles	Staged EU applicability (incl. GPAI obligations) creates time-bound demand [18]	Audit-ready "AI system of record": model inventory, usage logs, data provenance, incident response, and attestation workflows
Evaluation, observability, and reliability tooling	Falling inference cost increases experimentation and deployment; reliability becomes the limiter	Inference costs have fallen dramatically (example: GPT-3.5-level system cost from $20 to $0.07 per million tokens in ~18 months) [3], pushing competitive advantage from "can you run it?" to "can you measure and trust it?"	High if positioned as required for production (like APM)	Requires technical credibility; tough GTM without incumbency	Empirical evidence of rapidly increasing deployment scale and affordability [19]	Domain-specific eval harnesses; post-deployment "AI incident management"; agent reliability scoring tied to business KPIs

Shovel category	Why it's a "shovel" (wins regardless of model/app winner)	Demand / market proof points	Profitability potential	Entry barriers	Leading incumbency signals	Underserved opportunities still open
Data pipelines, quality, and rights management	Models are only as good as enterprise data access; rights + quality become gating	Central-bank governance guidance highlights data confidentiality and third-party dependencies as core risks of AI adoption [20]	Potentially very high due to switching costs	Needs connectors + governance + security	Regulatory and reputational sensitivity makes this non-optional for large orgs [21]	"Permissioned RAG" (row/column-level policy enforcement), data lineage for prompts/outputs, contractual "data nutrition labels"
Integration and workflow layers	Enterprises buy outcomes; integration turns generic models into usable systems	Financial-services evidence shows early genAI deployments cluster around content extraction, code assistance, risk/compliance, and internal ops—i.e., workflow integration [22]	High if embedded in core workflows; moderate if thin wrapper	Requires domain expertise + deep systems integration	Early use-case clustering suggests integration is where value is realized [23]	Vertical agent "connectors" to legacy systems; safe tool-use frameworks; migration kits from pilots to governed production

Shovel category	Why it's a "shovel" (wins regardless of model/app winner)	Demand / market proof points	Profitability potential	Entry barriers	Leading incumbency signals	Underserved opportunities still open
Education, enablement, and specialized services	Skills and implementation gaps persist; services monetize immediately and can productize later	Government and industry documents emphasize governance, training, and risk management as necessary complements to AI adoption [24]	Medium initially; can become high via productization	Low to moderate; reputation moat matters	Central banks explicitly recommend governance actions including training and tool inventories [20]	"Compliance-first AI implementation agencies; templates + controls + tooling bundles for SMEs in regulated sectors

AI Value Chain Dynamics and Where Profits Concentrate

This section maps the stack **Hardware → Models → Tools → Applications → Distribution → Services** while isolating (a) margin pools, (b) defensibility, and (c) saturation. The key meta-trend is that **inference is commoditizing faster than organizational risk tolerance is increasing**: model use gets cheaper and easier, while trustworthy deployment remains hard. [25]

Value chain map (margin, moat, crowding, and "underpriced" layers)

Layer	What's being sold	Typical profit logic	Most defensible when…	Crowding risk	"Underrated" wedge (what to build)
Hardware & components	GPUs/ accelerators, HBM, packaging, networking, servers	Scarcity + performance differentiation; supply chain rents	Capacity is constrained (HBM, packaging) and demand is rising	Low crowding at true chokepoints; high at commoditized servers	"Chokepoint operating systems": allocation, forecasting, and provenance across constrained components

Layer	What's being sold	Typical profit logic	Most defensible when…	Crowding risk	"Underrated" wedge (what to build)
Models & inference	Frontier foundation models, fine-tunes, inference APIs	Scale + distribution; but price pressure from falling inference costs	You have differentiated data + distribution + trust	Increasing; inference prices dropping sharply [3]	Multi-model routing, caching, governance, and "model-neutral" safety layers
Tools (LLMOps / eval / security)	Dev tooling, monitoring, security, governance	"Picks and shovels" inside the software factory; broad customer base	Tool becomes required control point (APM-like)	Medium-high; many startups	Regulated-domain eval and audit; agent permissioning tied to OWASP risks [15]
Applications	Vertical or horizontal AI products	Value-based pricing if ROI is provable	Deep workflow lock-in and proprietary distribution	Very high (thin wrappers) as models get cheaper [3]	"Boring automation" in regulated workflows where accuracy + audit matter
Distribution	Cloud marketplaces, enterprise suites, app ecosystems	Taxed access and privileged placement	You own the channel or default bundle	Structurally concentrated in big platforms [6]	Neutral distribution: procurement, billing, and compliance packaging for AI add-ons
Services	Implementation, training, managed operations	Monetize complexity immediately; can evolve to managed products	You become trusted operator-of-record	Competitive but local trust matters	Compliance-grade managed AI operations (outsourced "AI SOC" + "AI GRC")

Where margins are *actually* compressing vs compounding

Compressing:

- **Raw inference**: a concrete benchmark shows query costs collapsing (example: $20 → $0.07 per million

tokens for GPT-3.5-level performance in ~18 months), which tends to shift pricing power away from "model usage" toward "outcomes and trust." [3]

Compounding:
- **Bottlenecked compute inputs**: capacity constraints in advanced packaging and HBM support durable pricing power for chokepoint suppliers. [26]
- **Trust & regulatory compliance**: staged regulation (EU AI Act) and standardized risk frameworks (NIST AI RMF) institutionalize recurring spend on governance and control infrastructure. [27]
- **Energy + siting + interconnect**: data-center electricity demand is already significant and projected to rise sharply, expanding the profit pool for energy procurement, power management, and grid-adjacent software. [5]

Shovel-Seller Business Models and Moats

This section focuses on business-model choice as a function of **time-to-market**, **capital intensity**, and **defensibility**.

Which models are easiest to start now (and why)

Specialized services → productized services → SaaS is the dominant low-capital path in AI because the "unknown unknowns" in deployment are still high and buyers want accountability.

- **Agency / implementation studio (fastest cashflow):**
  Works because buyers face governance and risk-management needs *immediately* (inventory, policies, due diligence, monitoring). Central-bank governance guidance and US Treasury findings emphasize governance, risk management, and third-party diligence as core issues, which services can solve faster than pure software. [28]
- **Compliance-grade templates + managed operations:**
  Regulation creates repeatable checklists and artifacts. The EU AI Act's staged applicability provides a calendar of when different obligations begin to bite (eg, governance and general-purpose AI obligations earlier than full high-risk applicability). [9]
- **Security add-ons (prompt injection, data leakage, agent control):**
  OWASP's LLM Top 10 gives a shared taxonomy for buyers and sellers, which accelerates procurement because problems become legible. [14]

Which models are hardest—but could be most valuable long-term

- **Metered APIs at scale (model gateways, inference caches, evaluation-as-a-service):**
  Hard because platform incumbents can bundle, and pricing pressure is intense as inference costs fall. [3]
  Moat must come from *policy* + *routing intelligence* + *compliance evidence*, not from raw access to a model.
- **Marketplaces (data, tools, agents):**
  Hard because liquidity is difficult, and distribution is already concentrated in cloud ecosystems. [6]
  Winning marketplaces usually start as *workflow control points* (procurement, billing, compliance, or incident response), not as "listings sites."

- **Energy-adjacent infrastructure software:**
  Hard due to slow sales cycles and need for utility/permit interfaces—but the demand driver is clear and rising as data-center electricity grows. [13]

Practical moat checklist (what makes a shovel company durable)

A shovel seller becomes durable when at least one of these is true:

1. **Your product is a control point** (governance, security, metering, audit logging). NIST AI RMF and OWASP taxonomies turn "nice-to-have" controls into "must-have" controls. [29]
2. **You reduce a binding constraint** (power, compliance, integration, or scarce compute). Data-center electricity growth and supply-chain bottlenecks define binding constraints today. [30]
3. **You embed into a regulated workflow** (switching costs + audit trails). Financial-services guidance emphasizes explainability, third-party risk management, and governance requirements that naturally create switching costs. [31]

Outlook 2026–2030: What Grows Fast, What's Overhyped, What Becomes Infrastructure

Shovel businesses likely to grow fastest

Energy, power-management, and data-center enablement (including software)
Data-center electricity consumption is already estimated at ~415 TWh (~1.5% of global electricity) and is projected to grow dramatically over the next decade, surpassing 1,000 TWh by 2030 in a baseline case—implying multi-year capex and software spend waves around siting, interconnects, and operational optimization. [5]

Governance and compliance infrastructure
The EU AI Act's staged timeline (entered into force 2024; governance and GPAI obligations applying earlier than full applicability) strongly suggests 2026–2030 spend will concentrate on "auditability as a product feature," not just model performance. [9]
In parallel, standardized risk frameworks like NIST AI RMF and domain-specific governance guidance institutionalize risk-management demand. [32]

Security and fraud-defense upgrades
As AI increases both automation and adversarial capability, security spend rises. OWASP's taxonomy formalizes LLM-specific vulnerabilities and mitigations; in finance, regulators and law enforcement warn that criminals exploit generative AI to scale fraud and impersonation. [33]

What looks overhyped (as a "shovel")

Generic "LLM wrapper" apps without a control point
As inference becomes cheaper (large step-down shown in 18 months), thin-differentiation apps face margin compression and fast follower risk. The economics forces value capture to move to distribution, proprietary workflow hooks, and trust. [34]

One-model dependency as a strategy
The logic of shovel selling is model-agnostic: the more multi-model your customer base becomes, the more valuable routing, governance, and evaluation layers become. The same falling-cost trend incentivizes experimentation across models, reinforcing multi-model tool demand. [3]

What becomes "essential infrastructure"

1. **AI systems of record** (inventory, lineage, policy enforcement, audit logs) as default enterprise requirement, catalyzed by regulation schedules and governance frameworks. [35]
2. **Agent safety and permissioning** (tool-use constraints, secrets handling, and runtime policy checks), mapped directly to OWASP's LLM risk categories (prompt injection, insecure output handling, excessive agency). [36]
3. **Energy-aware compute orchestration** as grids tighten around major data-center clusters and electricity demand rises. [13]

Where individuals and small teams can still win

Small teams win where they can create leverage via **domain depth + distribution focus + compliance-grade trust**:

- **Regulated verticals** (finance, healthcare, govtech) where procurement prefers auditability and accountability. [27]
- **Integration-first wedges** into legacy systems, because value realization in early deployments often sits in workflow embedding and internal operations rather than "cool demos." [23]
- **Security+governance micro-products** that attach to existing stacks and expand later (the "control point first" play). [37]

Finance: AI Gold Rush Mapping and Shovel Opportunities

Finance gold rush mapping

Who are the "miners" in finance AI?
- Banks, broker-dealers, insurers, asset managers, and fintechs deploying AI directly into underwriting, trading, service, compliance, and operations. The US Treasury synthesis lists broad use across underwriting, trading/investment advice, compliance, forecasting, and process automation. [38]
- Market participants expanding model inventories and use cases as genAI adoption grows (the report notes expectations of significant model inventory increases among surveyed firms). [39]

Who are the shovel sellers in finance?
Finance "shovels" differ from general AI shovels in one major way: **regulatory and reputational risk are first-order**. Shovels therefore cluster around:

- **Fraud, identity, and financial crime controls**
- **Model governance and explainability artifacts** (especially for adverse actions and consumer impact)
- **Third-party risk and vendor oversight**
- **Secure data use and privacy**

These themes are explicit in US Treasury findings (third-party risk management, explainability, privacy, disclosure debates) and central-bank governance guidance (risk taxonomy including info security, third-party dependencies, model risks like hallucinations). [31]

Who controls infrastructure and distribution in finance?
- **Cloud + enterprise software ecosystems**: because many financial institutions increasingly adopt vendor-supported genAI tools for internal efficiency, distribution is mediated through vendors and enterprise procurement. [40]
- **Regulators and self-regulators** define what "acceptable" looks like: for example, US regulators emphasize explainability obligations in lending decisions even when complex models are used. [41]
- **Global AML/CFT standard setters** influence fraud tooling requirements: FATF explicitly discusses AI-enabled deepfake risks and the need for strengthened safeguards. [42]

Finance shovel opportunity map by category

The most reliable "finance shovels" monetize one of three budgets: **(1) fraud losses avoided, (2) compliance cost reduced, (3) revenue uplift with auditable controls**.

Category	Problem being solved	Who pays	Revenue potential	Competition level	Ease of entry	Future demand drivers
Risk, fraud, identity, account takeover	AI raises attacker capability (voice/video impersonation, synthetic IDs); institutions must detect, verify, and respond	Banks, brokers, payment providers, fintechs	Very high (fraud is a direct P&L line); can be usage-based per check or per account	High, but still expanding	Moderate (needs high-quality detection + integrations)	FBI warns AI is used for convincing impersonation; FINRA warns of genAI-enabled account fraud; FATF flags AI/ deepfake AML risks [43]
AML/KYC and financial crime ops ("agentic compliance")	Manual onboarding, periodic reviews, alert triage; need faster, event-driven diligence	Banks and regulated fintechs	High ACV if you reduce headcount/ time per case	Medium-high	Moderate (domain + integration heavy)	McKinsey describes agentic AI automating end-to-end KYC workflows; FATF is explicitly focused on AI/ deepfakes risk in AML context [44]

Category	Problem being solved	Who pays	Revenue potential	Competition level	Ease of entry	Future demand drivers
Credit underwriting & explainability	Complex models create “black box” issues; regulators require specific reasons for adverse actions	Lenders, BNPL, credit unions, core platforms	High in B2B; sticky if embedded	Medium	Moderate-high (needs legal + data + model interpretability)	CFPB guidance stresses specific, accurate adverse action reasons when AI/complex models are used [45]
Trading & execution tooling	AI-assisted research, signal generation, execution optimization; governance and model risk remain	Hedge funds, prop shops, banks	High but concentrated; winner-take-most at top	High	Hard (needs differentiated edge + data)	Treasury notes AI is widely used in investment/ trading; but governance and risk concerns persist [46]
Compliance/ regtech for communications and surveillance	Large volumes of comms and trades; need automated review + audit trails	Broker-dealers, banks	High recurring SaaS	Medium-high	Moderate	FINRA oversight reports increasingly discuss genAI risks and the need for controls; rising scrutiny increases tooling spend [47]

Category	Problem being solved	Who pays	Revenue potential	Competition level	Ease of entry	Future demand drivers
Model risk management (MRM) and third-party AI risk	Vendor models + internal models create concentration and third-party dependency; requires inventories, due diligence, monitoring	Banks, insurers, large fintechs	High ACV; sticky “system of record”	Emerging but increasingly crowded	Moderate (requires trust + frameworks)	Treasury highlights concentration and third-party risk management needs; BIS central-bank guidance stresses third-party and model risks [31]
Wealth management and advisor copilots	Productivity tools for advisors (research, summaries, suitability) with compliance controls	Wealth platforms, RIAs, broker-dealers	Medium-high; depends on distribution	Medium	Moderate	Firms proceed cautiously with vendor-supported genAI tools for internal efficiency; compliance scrutiny remains [48]
Finance ops / CFO tooling (close, reconciliation, narratives)	Automate close, reconciliations, narrative reporting; reduce manual errors	Mid-market companies, CFO orgs, accounting firms	High volume SMB/mid-market	Medium-high	Easier (clear workflows)	Treasury notes process automation and content extraction are early integration areas [23]
Personal finance AI (B2C)	Guidance, budgeting, negotiation; but trust and liability are hard	Consumers (often indirectly via partner rev)	Mixed; CAC heavy	Very high	Easier to build, hard to scale	Consumer trust and fraud risks intensify; compliance and liability considerations matter [49]

Category	Problem being solved	Who pays	Revenue potential	Competition level	Ease of entry	Future demand drivers
Data providers and permissioned data layers	Securely use sensitive financial data in AI systems without leakage; enable "permissioned RAG"	Institutions with large data estates	High if embedded	Medium	Hard (integrations + security)	Central banks emphasize confidentiality risk; AI RMF emphasizes lifecycle risk management; EU AI Act increases compliance drive [50]

Actionable Opportunity Stack and Positioning Strategy

Ten best AI shovel opportunities overall

These are ranked by a VC/operator blend: **non-discretionary demand + defensibility + timing + ability for small teams to wedge in**.

1. **AI Governance System of Record (inventory → policy → audit logs → attestations)**
   Demand is pulled forward by regulatory timelines and standardized risk expectations. [27]
2. **Agent Permissioning + Runtime Safety Layer (tool-use control, secrets, data boundaries)**
   Maps directly to OWASP risk categories like prompt injection, insecure output handling, and excessive agency. [36]
3. **LLM/Agent Evaluation Harness for Regulated Workflows (finance/health/gov)**
   Inference gets cheap; trust gets expensive—so "measurement" becomes the bottleneck. [3]
4. **Energy-aware AI Compute Orchestration (carbon- and grid-aware routing, demand response)**
   Data-center electricity growth makes this an infrastructure problem, not a niche feature. [51]
5. **Multi-cloud AI FinOps (token/unit economics, caching, routing, budget controls)**
   The faster inference costs fall, the more usage can explode—making governance and cost control essential. [3]
6. **Supply-chain Provenance and Compliance for AI Components (model + data + compute)**
   Capacity constraints and third-party dependencies increase the value of provenance and risk controls. [52]
7. **Enterprise "Permissioned RAG" Data Layer (policy enforcement + lineage)**
   Confidentiality and privacy risks are repeatedly highlighted in governance guidance. [28]
8. **Security Testing / Red Teaming for LLMs and Agents (continuous, benchmarked)**
   OWASP provides a shared language for procurement; enterprises need repeatable testing. [15]
9. **Vertical Integration Kit: "Pilot → Production" in a regulated niche**
   Early adoption clusters around content extraction and risk/compliance; packaging the path to production is valuable. [23]

10. **Education-to-Product funnel: compliance-grade enablement + tooling bundles**
    Governance bodies emphasize training, tools inventories, and risk processes; education is a wedge into recurring tooling. [24]

Ten best AI shovel opportunities in finance

1. **Deepfake-resistant identity verification and step-up authentication**
   Law enforcement explicitly warns of AI-powered impersonation; finance must respond. [53]
2. **GenAI-native fraud ops platform (case triage + evidence + reporting)**
   FINRA warns fraudsters use genAI to create new accounts and take over accounts; tooling that shortens time-to-detection and improves evidence capture sells. [54]
3. **Agentic KYC/AML workflow automation with audit trails**
   McKinsey describes end-to-end KYC automation potential; FATF flags AI-enabled fraud modalities. [44]
4. **Adverse action explainability tooling for AI/complex underwriting models**
   CFPB guidance stresses specific, accurate reasons for credit denials even when AI is used. [41]
5. **Third-party AI risk management platform for regulated firms (contracts → controls → monitoring)**
   Treasury highlights concentration and third-party risk management needs; BIS governance guidance echoes external dependency risks. [31]
6. **Model risk management modernization: "AI model inventory + continuous monitoring"**
   As model inventories expand, governance and monitoring become recurring spend. [55]
7. **Surveillance and supervision tooling that detects AI-generated market manipulation content**
   FINRA materials increasingly highlight genAI-related risks; supervision budgets follow scrutiny. [56]
8. **Secure data access layer for genAI in banks (policy enforcement + confidential RAG)**
   Central-bank guidance emphasizes confidentiality and security risks; solving safe data access is core. [20]
9. **Compliance-grade advisor copilot (suitability + documentation + disclosures)**
   Firms are cautious and vendor-supported tools are common; packaging controls is the wedge. [57]
10. **Financial crime threat intel for AI-enabled scams (playbooks + detection rules)**
    FBI and FINRA both emphasize AI-enabled fraud threats, creating demand for specialized intelligence and controls. [58]

Skills to learn now to benefit (high leverage, shovel-aligned)

- **Evaluation and measurement of model/agent behavior** (business-metric-tied evals, failure taxonomy), because inference affordability shifts the bottleneck to reliability and trust. [3]
- **AI security fundamentals** grounded in OWASP's LLM risk taxonomy (prompt injection, data leakage, agent misuse, model DoS). [14]
- **AI governance and risk artifacts** (model inventories, audit logs, controls mapped to NIST AI RMF and relevant regulation). [59]
- **Data engineering for permissioned data access** (lineage, policy enforcement, privacy-by-design), reflecting confidentiality and privacy risks emphasized by financial authorities. [28]
- **Finance domain depth in one workflow** (AML/KYC, underwriting, surveillance, treasury ops), because the US Treasury evidence suggests early genAI integration is function-specific, not generic. [23]

Positioning strategy for someone entering today

Position as the "trust layer," not the "model layer."
The empirical direction of the market is: models get cheaper and more available, while the cost of mistakes (security incidents, regulatory violations, reputational damage) rises. Inference cost declines (order-of-magnitude drops) imply that differentiation shifts to governance, security, integration, and measurable outcomes. [60]

A practical entry strategy that repeatedly works in "gold rush" markets:

1. Pick **one regulated workflow** with clear pain and budgets (fraud ops, KYC onboarding, adverse action notices). [61]
2. Start with a **service that produces compliance-ready artifacts** (policies, logs, evaluation reports), aligned to established frameworks. [32]
3. Productize the repeatable layer into the **system of record / control point**, where switching costs form naturally.

If AI is a gold rush, the smartest move today is: build the control points—governance, security, and workflow-integrated infrastructure—that every AI "miner" must rely on as compute scales and regulation tightens.

---

[1] Economic potential of generative AI
https://www.mckinsey.com/capabilities/tech-and-ai/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier?utm_source=chatgpt.com

[2] [3] [19] [25] [34] [60] https://hai.stanford.edu/assets/files/hai_ai-index-report-2025_chapter1_final.pdf
https://hai.stanford.edu/assets/files/hai_ai-index-report-2025_chapter1_final.pdf

[4] [8] [11] [26] [52] https://www.reuters.com/technology/tsmc-considering-advanced-chip-packaging-capacity-japan-sources-say-2024-03-17/
https://www.reuters.com/technology/tsmc-considering-advanced-chip-packaging-capacity-japan-sources-say-2024-03-17/

[5] [30] Energy demand from AI
https://www.iea.org/reports/energy-and-ai/energy-demand-from-ai?utm_source=chatgpt.com

[6] Cloud Market Share Trends - Big Three Together Hold 63 ...
https://www.srgresearch.com/articles/cloud-market-share-trends-big-three-together-hold-63-while-oracle-and-the-neoclouds-inch-higher?utm_source=chatgpt.com

[7] [14] [15] [16] [33] [37] https://owasp.org/www-project-top-10-for-large-language-model-applications/
https://owasp.org/www-project-top-10-for-large-language-model-applications/

[9] [18] [27] [35] AI Act | Shaping Europe's digital future - European Union
https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai?utm_source=chatgpt.com

[10] https://www.nps.gov/podcasts/better-lives-bitter-lies.htm?sortby=date-asc
https://www.nps.gov/podcasts/better-lives-bitter-lies.htm?sortby=date-asc

[12] CFO Commentary on Fourth Quarter and Fiscal 2025 Results
https://www.sec.gov/Archives/edgar/data/1045810/000104581025000021/q4fy25cfocommentary.htm?utm_source=chatgpt.com

13 51 Energy supply for AI
https://www.iea.org/reports/energy-and-ai/energy-supply-for-ai?utm_source=chatgpt.com

17 https://www.nist.gov/itl/ai-risk-management-framework
https://www.nist.gov/itl/ai-risk-management-framework

20 21 24 28 50 https://www.bis.org/publ/othp90.pdf
https://www.bis.org/publ/othp90.pdf

22 23 31 38 39 46 49 55 https://home.treasury.gov/system/files/136/Artificial-Intelligence-in-Financial-Services.pdf
https://home.treasury.gov/system/files/136/Artificial-Intelligence-in-Financial-Services.pdf

29 32 59 https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf
https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf

36 https://www.cloudflare.com/learning/ai/owasp-top-10-risks-for-llms/
https://www.cloudflare.com/learning/ai/owasp-top-10-risks-for-llms/

40 47 48 57 https://www.finra.org/media-center/newsreleases/2025/finra-publishes-2025-regulatory-oversight-report
https://www.finra.org/media-center/newsreleases/2025/finra-publishes-2025-regulatory-oversight-report

41 45 61 https://www.consumerfinance.gov/about-us/newsroom/cfpb-issues-guidance-on-credit-denials-by-lenders-using-artificial-intelligence/
https://www.consumerfinance.gov/about-us/newsroom/cfpb-issues-guidance-on-credit-denials-by-lenders-using-artificial-intelligence/

42 https://www.fatf-gafi.org/en/publications/Methodsandtrends/horizon-scan-ai-deepfake.html
https://www.fatf-gafi.org/en/publications/Methodsandtrends/horizon-scan-ai-deepfake.html

43 53 https://www.fbi.gov/contact-us/field-offices/sanfrancisco/news/fbi-warns-of-increasing-threat-of-cyber-criminals-utilizing-artificial-intelligence
https://www.fbi.gov/contact-us/field-offices/sanfrancisco/news/fbi-warns-of-increasing-threat-of-cyber-criminals-utilizing-artificial-intelligence

44 How agentic AI can change the way banks fight financial ...
https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/how-agentic-ai-can-change-the-way-banks-fight-financial-crime?utm_source=chatgpt.com

54 https://www.finra.org/investors/insights/gen-ai-fraud-new-accounts-and-takeovers
https://www.finra.org/investors/insights/gen-ai-fraud-new-accounts-and-takeovers

56 https://www.finra.org/sites/default/files/2025-01/2025-annual-regulatory-oversight-report.pdf
https://www.finra.org/sites/default/files/2025-01/2025-annual-regulatory-oversight-report.pdf

58 https://www.ic3.gov/PSA/2024/PSA241203
https://www.ic3.gov/PSA/2024/PSA241203