BOOK FOUR-B · THE AI ECONOMY MONETIZATION SERIES

The AI Marketplace Economy

How AI agents are discovered, priced, and monetized — and how to build the platforms that power it

In the B2B AI economy, discovery is destiny. The marketplace that surfaces the right capability at the right moment — to the right enterprise buyer or AI agent — owns the commercial gravity of the entire ecosystem.

Five layers. Three B2B participant types. One governing principle: trust enables commerce, and commerce compounds trust.

Framework F24: The AI Marketplace Architecture Model · Infrastructure Pair: Books 4 and 4b

PREFACE

From eBay to the AI Economy

How the oldest idea in commerce — the marketplace — becomes the most important infrastructure layer in the AI economy.

In 2001, when the first enterprise software procurement portals appeared — early predecessors to what would become AWS Marketplace, Salesforce AppExchange, and ServiceNow Store — the promise was simple: reduce the friction of finding, evaluating, and purchasing enterprise software. The enterprise buyer's problem was not shortage of supply; it was abundance of supply with insufficient discovery and evaluation infrastructure. The vendor's problem was not shortage of capability; it was the enormous

cost of building the trust credentials, the sales relationships, and the procurement integration required to reach enterprise buyers at scale.

Two decades later, the same structural problems appear in the B2B AI capability market — but at dramatically higher velocity, dramatically greater complexity, and with an entirely new buyer type that did not exist in 2001: the AI agent itself.

The B2B AI marketplace is the infrastructure that solves these problems simultaneously. For the enterprise AI team evaluating a legal analysis AI, a clinical documentation AI, or a financial forecasting model, the marketplace provides structured discovery (what is available, how does it compare, what has been independently benchmarked), trust evidence (security certifications, compliance documentation, domain-specific performance attestation), and commercial infrastructure (standard pricing, procurement integration, contract templates). For the AI capability vendor — whether a specialized model creator, an AI application company, or a foundation model provider — the marketplace provides distribution reach, trust certification, and billing infrastructure that would take years and tens of millions of dollars to replicate independently.

This book covers both sides of that value exchange — but it covers them entirely within the B2B context. There are no consumers here. There are no individual hobbyists, no personal subscriptions, no retail commerce in the consumer sense. Every buyer in this book is a business: an engineering organization, a procurement team, an IT department, a data science function, or — increasingly — an AI agent operating with commercial authority on behalf of a business.

One terminology clarification before proceeding, because it prevents significant confusion throughout the book: when we use the word "retail" in reference to a B2B customer or a B2B channel, we mean the online-channel-originated B2B customer — the company that discovers and purchases through a self-serve digital interface rather than through a direct sales engagement. A technology startup that signs up for Hugging Face Pro through the website is a retail B2B customer. A Fortune 500 that negotiates a custom enterprise agreement through AWS Marketplace Private Offers is an enterprise

B2B customer. Both are B2B. The retail/enterprise distinction is a channel and contract-complexity distinction, not a business-versus-consumer distinction.

The B2B AI marketplace is at an inflection point in 2025. The capability supply is abundant but under-discoverable. The enterprise buyer demand is accelerating but under-served by current discovery and trust infrastructure. The agent-to-agent commerce infrastructure that will eventually dominate marketplace transaction volume is being built right now, in the developer tools, research, and data analytics verticals. And the competitive consolidation that will determine which platforms own the dominant B2B AI marketplace positions is in progress — the window to establish the network effects, trust credentials, and commercial infrastructure that create durable competitive positions is open but narrowing.

This book is the playbook for operating in this moment — as a platform builder, as a B2B AI capability vendor, or as an enterprise buyer trying to navigate an increasingly complex procurement landscape with confidence.

IMPORTANT TERMINOLOGY: RETAIL B2B vs CONSUMER

In this book, retail means online-channel B2B — not consumer

Before proceeding, one definitional matter demands explicit treatment because the terminology causes genuine commercial confusion. This book uses the word "retail" in its B2B sense: a B2B customer who discovers, evaluates, and purchases an AI capability through a self-serve online channel — a marketplace listing, a developer portal, a PLG funnel, or a SaaS sign-up — is a retail B2B customer. "Retail" here means online-channel-originated, not consumer-facing. The distinction between a retail B2B customer and an enterprise B2B customer in the AI marketplace context: A retail B2B customer is a business buyer — typically an engineering team, a data science organization, or a product team at a company — who discovers the AI capability through the marketplace's online discovery infrastructure, evaluates it through the marketplace's self-serve tooling, and purchases it through the marketplace's standard commercial terms without a dedicated sales engagement. The transaction is commercial and B2B in nature; the channel is digital and self-serve. Hugging Face's Pro and Team tier customers, Replicate's API subscribers, and AWS Marketplace standard-listing customers are all retail B2B customers by this definition. An enterprise B2B customer is also a business buyer

— typically an IT organization, a procurement team, or a C-suite-sponsored initiative at a large organization — who discovers the AI capability through a combination of marketplace research and direct vendor engagement, evaluates it through a formal procurement process that may include security review, compliance assessment, and contract negotiation, and purchases it through a custom commercial agreement (Private Offer, Enterprise Agreement, or negotiated MSA). The transaction is commercial and B2B; the channel is direct or semi-direct, with the marketplace providing discovery infrastructure and trust signals even when the final transaction occurs outside the marketplace's standard flow. Both are B2B customers. Neither is a consumer. The word "retail" in this book always means the online-channel-originated B2B customer — the company that signs up through the website rather than through a sales team. This distinction matters commercially because retail B2B customers and enterprise B2B customers require different trust infrastructure, different pricing models, different contract structures, and different commercial motions — but both are served by the same AI marketplace platform, entering through different channels.

PART ONE

The B2B AI Marketplace Landscape

Three archetypes. Why now. The economics of discovery.

CHAPTER ONE

The Three Marketplace Archetypes

Model and dataset hubs · Enterprise application marketplaces · Cloud AI service markets. Three structures, three commercial logics.

The B2B AI marketplace landscape has three structurally distinct archetypes, each serving different B2B participant types, monetizing differently, and competing on different dimensions. The archetype determines every significant platform design and commercial decision.

The Developer and ML Practitioner Hub is the first archetype. Its supply participants are ML researchers, academic groups, AI startups, and foundation model providers who publish models, datasets, and tools. Its demand participants are B2B engineering and data science organizations — companies that embed AI capabilities into their own products and internal systems. The retail B2B buyer at this archetype is a developer team that discovers and subscribes to AI capabilities through a self-serve online interface: a Hugging Face Pro subscription, a Replicate API key, a fal.ai account. The enterprise B2B buyer at this archetype is the organization that negotiates a private model hosting agreement, a dedicated inference endpoint contract, or an Enterprise Hub arrangement. Both are B2B. Both are served by the same platform with different commercial motions.

Hugging Face dominates this archetype. With over one million public models and a developer community of millions, Hugging Face has achieved the critical mass at which the network effects are self-sustaining: new models are published there reflexively, because that is where the audience is; enterprise buyers discover there because that is where the supply is; the community evaluation infrastructure (Open LLM Leaderboard, community model reviews) creates the trust signals that allow retail B2B buyers to make informed purchasing decisions without the extended evaluation cycles that enterprise procurement requires.

The Enterprise Application Marketplace is the second archetype. It is embedded within a larger enterprise software platform — Salesforce, ServiceNow, Atlassian, SAP, Workday — and serves the enterprise buyers who are already committed to that platform's ecosystem. The supply participants are ISVs and AI application companies who build products that extend the platform's core capabilities. The demand participants are enterprise IT teams and business operations groups who need AI capabilities that integrate with their existing enterprise systems.

There is no "retail B2B" buyer at this archetype in the sense of individual developer sign-ups — the minimum viable buyer is an enterprise organization with an existing platform relationship. However, within an enterprise organization, the channel dynamic is

similar: some application purchases go through IT-led enterprise procurement (enterprise B2B channel), while others go through departmental self-service purchases by business users who discover applications in the marketplace's app store interface (retail B2B channel within the enterprise). Both channels serve the same enterprise organization; the distinction is in the procurement authority and process.

Salesforce AppExchange with its 7,000+ applications and $5B+ in cumulative transaction volume is the most mature example of this archetype. The AI extension — AI-powered applications built on foundation models that extend Salesforce's core capabilities — is among the fastest-growing categories on AppExchange. The trust infrastructure Salesforce built (the Security Review process, the partner certification program) is the template that AI application marketplaces in every platform ecosystem are replicating.

The Cloud AI Service Marketplace is the third archetype. It lives inside a cloud provider's commercial ecosystem — AWS, Azure, GCP — and serves enterprise buyers who are already spending significantly with that cloud provider. The supply participants are AI companies that have achieved the certification and commercial readiness required for cloud marketplace listing. The demand participants are enterprise technology organizations whose procurement relationship with the cloud provider creates the single-invoice, credits-applicable purchasing simplicity that drives marketplace adoption.

The retail B2B channel at this archetype is the standard marketplace listing: an enterprise developer or technology team discovers a listed AI service, subscribes through the standard marketplace interface, and the subscription appears on the organization's monthly cloud bill. The enterprise B2B channel is the Private Offer or Channel Partner Private Offer (CPPO) structure: custom commercial terms negotiated between the AI vendor and the specific enterprise customer, executed through the cloud marketplace interface but with pricing and terms that are specific to that enterprise relationship.

AWS Marketplace's AI and Machine Learning category has grown significantly as enterprise buyers discovered that standard AI service subscriptions could be added to

their existing AWS relationship without new vendor onboarding. The competitive limitation of all cloud provider marketplaces is their cloud-locked nature: the marketplace's procurement integration moat only applies to buyers standardized on that cloud provider's infrastructure.

ARCHETYPE 1: MODEL AND DATASET HUB *Developer-facing infrastructure. Community flywheels. Open-source distribution meets commercial infrastructure.*	
Primary participants	ML practitioners, researchers, and AI developers (supply) · AI developers, data scientists, and enterprises (demand)
Primary assets	Pre-trained models, fine-tuned models, datasets, evaluation benchmarks, inference tooling, and demo applications (Spaces)
Monetization model	Free community hosting (builds network effects) + paid infrastructure (Inference API, private repos, Enterprise Hub) + developer tooling subscriptions
Trust mechanism	Community-driven evaluation benchmarks (Open LLM Leaderboard), model cards, usage counts, community reviews, automated safety scans
Network effect type	Data network effect (usage data improves recommendations) + supply-demand flywheel + integration network (Transformers library makes every model accessible)
Winner	Hugging Face — one million+ public models, first-mover advantage, Transformers library integration creates structural platform lock
Competitive dynamics	Near-monopoly in ML model discovery; challenged in inference by Replicate/fal.ai; challenged in enterprise by cloud provider marketplaces

ARCHETYPE 2: ENTERPRISE APPLICATION MARKETPLACE *Platform-adjacent app stores. Trust as the primary moat. Take-rate economics at enterprise scale.*	
Primary participants	ISVs and AI application companies (supply) · Enterprise IT teams and business users in the platform's installed base (demand)
Primary assets	SaaS applications extending a platform's core functionality: AI-powered CRM tools, workflow automations, industry-specific AI applications
Monetization model	Take rate on app subscription revenue (15–25%) + listing and certification fees + enterprise co-sell commission structures
Trust mechanism	Security review certification (AppExchange security badge), compliance

	certifications, partner tier system, enterprise customer reviews
Network effect type	Integration network effect (each certified integration creates switching costs) + supply-demand flywheel within the platform ecosystem
Winner(s)	Salesforce AppExchange (CRM), Shopify App Store (e-commerce), Atlassian Marketplace (dev tools), ServiceNow Store (enterprise workflow)
Competitive dynamics	Winner-takes-most within each platform ecosystem; cross-platform competitive dynamics limited by platform lock-in

ARCHETYPE 3: CLOUD AI SERVICE MARKETPLACE *Procurement-integrated distribution. Cloud-native buyers. Revenue-share economics at cloud scale.*	
Primary participants	AI companies seeking enterprise distribution (supply) · Enterprise IT teams already in a cloud provider's ecosystem (demand)
Primary assets	Managed AI services: inference APIs, specialized AI tools available as cloud services, AI-powered data processing, AI safety and governance services
Monetization model	Revenue share between cloud provider and seller (3–25% depending on service type) + CPPO (private offer) structures for custom enterprise terms
Trust mechanism	Cloud provider security review, compliance certifications inherited from cloud provider's compliance posture, private offer structure for enterprise-specific terms
Network effect type	Procurement integration network effect (cloud spending pools) + supply-demand flywheel within cloud provider ecosystem
Winner(s)	AWS Marketplace (largest), Azure Marketplace, GCP Marketplace; AWS Marketplace dominant by transaction volume
Competitive dynamics	Each cloud marketplace serves its own ecosystem; cross-cloud neutral marketplaces (not yet mature) represent the most significant whitespace

Chapter One — The Essentials

› Three distinct archetypes: model/dataset hub (developer-facing, community flywheel), enterprise app marketplace (platform-adjacent, trust-gated), cloud AI service marketplace (procurement-integrated, cloud-native).

› Each archetype serves different participant types, monetizes differently, and competes on different dimensions — the archetype determines every significant design and commercial decision.

› Hugging Face dominates the model hub archetype by a wide margin; the enterprise and cloud

archetypes have multiple strong incumbents in each platform ecosystem.

› The most significant whitespace: a cloud-neutral enterprise AI marketplace that provides the discovery, trust, and compliance infrastructure of cloud provider marketplaces without the cloud-specific procurement integration.

› Understanding which archetype you are building — or participating in — determines your growth strategy, your commercial model, and your competitive moat.

CHAPTER TWO

Why Now: Four Converging Forces

The capability explosion. The distribution cost collapse. The enterprise build-to-buy shift. Agent commerce.

The B2B AI marketplace economy is reaching an inflection point in 2025 for four converging reasons that together create conditions for rapid value creation and competitive consolidation.

The enterprise AI procurement bottleneck has become severe. Enterprise technology organizations are evaluating AI capabilities at a pace that their traditional procurement processes cannot support. A large bank evaluating AI for regulatory compliance monitoring may need to assess dozens of relevant models and services — each requiring vendor onboarding, security review, compliance assessment, and contract negotiation under traditional procurement processes. Without marketplace infrastructure that pre-clears trust requirements and standardizes commercial terms, the evaluation and procurement cycle consumes months that enterprise AI teams do not have. The B2B AI marketplace that builds comprehensive trust pre-clearance is worth a meaningful procurement cycle reduction to every enterprise buyer it serves.

The distribution cost for specialized B2B AI vendors has become prohibitive. A company that builds a specialized AI for protein structure prediction, legal contract analysis, or supply chain optimization faces a genuine distribution problem: reaching the enterprise buyers who would pay meaningful prices for the capability requires enterprise sales

teams, compliance certifications, procurement integrations, and legal capabilities that cost more to build than the initial revenue justifies. The B2B AI marketplace eliminates most of this distribution overhead — the vendor lists, certifies, and makes their capability available to the marketplace's enterprise buyer base without building a standalone commercial infrastructure organization.

Enterprise technology organizations are shifting from build to buy for specialized AI components. As discussed in Book 7, enterprises are increasingly capable of building broad AI capabilities internally using AI coding tools. But this internal build capability actually increases demand for specialized, best-of-breed AI components that are too narrow to justify internal development. A company building an internal AI assistant for its legal team does not want to also build a specialized contract analysis model, a regulatory compliance checker, and a legal citation verification engine — it wants to compose those capabilities from the B2B AI marketplace. The more enterprises build, the more they need to buy specific components.

Agent-to-agent commerce is already operating in B2B contexts. In software development, research, and data analytics verticals, AI orchestration agents are already invoking specialized AI capabilities through API calls that are economically equivalent to marketplace transactions — even when they are technically direct integrations rather than true marketplace commerce. The commercial infrastructure that makes these de facto transactions into true B2B marketplace transactions (dynamic capability discovery, real-time pricing, micropayment settlement, agent identity governance) is the next layer being built. The B2B AI marketplace that builds this infrastructure first captures the automated transaction volume that will eventually exceed human-initiated marketplace transactions.

The Four Converging Forces Driving AI Marketplace Growth			
Force	**What it creates**	**Evidence**	**Commercial implication**
Capability explosion outpacing	Buyers cannot evaluate the available options	Healthcare organizations face hundreds of relevant AI options for any given	Marketplace that provides structured discovery captures

discovery	without marketplace infrastructure	clinical workflow; evaluation without marketplace tools takes months	enterprise buyers who cannot navigate the unmediated landscape
Distribution cost collapse for specialized creators	Marketplace becomes the viable commercial path for specialized model creators who cannot build direct commercial infrastructure	Research teams at universities, independent AI researchers, and small AI companies now publish commercially viable models; without marketplace, they would have no path to revenue	Provider supply grows as distribution barriers fall; more supply strengthens the network effect for buyers
Enterprise shift from build to buy for narrow capabilities	Enterprises building broad AI internally need specialized AI capabilities they cannot justify building themselves	Companies building internal coding assistants still need security scanners, test generators, documentation tools — all marketplace opportunities	Enterprise buyer demand grows as the build decision creates more specific purchasing needs for components
Agent-to-agent commerce emerging in specific verticals	Automated marketplace transactions at machine speed are already operating in coding, research, and data analysis workflows	LangChain tool libraries, CrewAI agent networks, and Anthropic's Claude tool use are executing de facto marketplace transactions at increasing volume	A2A commerce infrastructure (micropayment settlement, real-time pricing APIs, agent identity) creates a new transaction category that human-speed procurement cannot serve

> ***"The AI capability explosion has created a discovery problem. The marketplace that solves it first — with sufficient trust infrastructure for enterprise deployment — captures the most valuable distribution position in the AI economy."***

Chapter Two — The Essentials

› Four converging forces create a structural marketplace opportunity: capability explosion, distribution cost collapse, enterprise build-to-buy shift, and agent commerce emergence.

› Each force independently justifies marketplace investment; their combination creates a specific window of competitive opportunity before market consolidation.

› The A2A commerce force is the most forward-looking: it will generate more transaction volume than human-mediated marketplace transactions within five years.

› The timing asymmetry: discovery infrastructure and trust infrastructure take 3–5 years to build to enterprise-grade quality — the time to start is before the buyer demand peaks.

› The first marketplace to establish critical mass in a specific vertical will be extraordinarily difficult to displace — the network effect advantage compounds faster than any competitor can match.

CHAPTER THREE

B2B Customer Segments: Retail, Enterprise, Departmental, and Agent

What each participant type needs, what they get, and how the take rate must reflect real value added.

Understanding the four B2B customer segments that a B2B AI marketplace serves is essential for designing the right discovery, trust, and commercial infrastructure for each. These segments differ in procurement authority, evaluation sophistication, budget size, and commercial relationship requirements — but all four are B2B customers, none are consumers.

The retail B2B customer (online channel, individual organization account) is a business — an engineering team at a startup, a data science group at a mid-market company, or a product team at a technology company — that discovers and purchases AI capabilities through the marketplace's self-serve digital interface. The retail B2B customer signs up with a credit card or a company purchase order, subscribes to a standard commercial plan, and is governed by the marketplace's standard terms of service. The commercial

relationship is standardized. The evaluation process is self-directed. The sales motion is product-led. The pricing is catalog pricing.

The retail B2B segment is commercially important because of its volume: the majority of marketplace transactions by count originate from retail B2B customers. It is also important because of its growth trajectory: retail B2B customers often graduate to enterprise B2B relationships as their AI deployment matures and their commercial requirements become more complex. The Hugging Face Pro subscriber who starts at $9/month and grows into an Enterprise Hub customer at $50K/year is the prototypical retail-to-enterprise B2B journey.

The enterprise B2B customer (direct channel, negotiated terms) is a large organization — typically with 1,000+ employees, a dedicated IT organization, and formal technology procurement processes — that purchases AI capabilities through a combination of marketplace discovery and direct commercial negotiation. The enterprise B2B customer requires custom contract terms (MSA, DPA, SLA), compliance documentation (SOC 2, HIPAA, FedRAMP as applicable), security review clearance, and procurement process integration (PO-based billing, ERP integration, vendor risk assessment). The commercial relationship is customized. The evaluation process is formal. The sales motion is enterprise-assisted. The pricing is negotiated.

The enterprise B2B segment is commercially important because of its value: the majority of marketplace transaction value originates from enterprise B2B customers. A single enterprise contract may be worth as much as hundreds of retail B2B subscriptions. The marketplace's commercial infrastructure must support both the standard-terms retail relationship and the negotiated-terms enterprise relationship within the same platform.

The departmental B2B customer (hybrid channel, departmental authority) is a business unit or department within a larger enterprise organization that has been granted limited procurement authority — typically up to a configured annual threshold — to purchase AI capabilities without full IT-led enterprise procurement. The departmental buyer discovers capabilities through the marketplace's online interface (retail channel behavior) but purchases on behalf of a larger organization with organizational-level

trust credentials already established. In enterprise technology organizations that have adopted cloud infrastructure, departmental purchasing under an enterprise cloud account is the norm for AI service subscriptions below the enterprise procurement threshold.

The agent-buyer (automated channel, programmatic authority) is an AI agent operating on behalf of a business organization, purchasing AI capabilities autonomously within its configured commercial authority. The agent-buyer is strictly B2B: it is acting as an agent of a business organization, not as an individual. The commercial authority it exercises has been delegated to it by a human within the business organization. Every transaction it executes is attributable to, and charged to, the business organization it represents. The commercial infrastructure for agent buyers — identity, authority limits, real-time pricing APIs, micropayment settlement — is B2B infrastructure, not consumer infrastructure.

Participant Economics — Value Exchange Reference				
Participant	**What they need**	**What the marketplace provides**	**Alternative without marketplace**	**Value retained vs captured**
Model/capability providers	Distribution to buyers they cannot reach directly; trust credibility that enterprise buyers require; infrastructure they would otherwise build	Buyer audience (discovery reach), certification that unlocks enterprise procurement, billing/API/compute infrastructure, usage analytics and market intelligence	Direct distribution: possible for large providers; prohibitive for specialized creators	Provider retains IP and revenue minus take rate; gains distribution they could not generate at comparable cost
Enterprise buyers	Efficient discovery of relevant options; reliable evaluation enabling	Taxonomy and search, benchmark data and model cards, security certification and compliance documentation,	Direct evaluation: each vendor requires individual security review	Buyer retains 100% of capability value; pays marketplace premium for discovery

	comparison; trust allowing production deployment; procurement that fits existing workflows	cloud billing integration or simplified vendor onboarding	(months); no standardized comparison; significant procurement overhead	efficiency and trust infrastructure
The marketplace platform	Take rate high enough to sustain infrastructure; network large enough to generate compounding data advantages; incentive structures that retain participants	Creates the trust and discovery infrastructure that enables all participant value; captures network data that improves all participants' experience over time	No comparable position available — marketplace is the category that connects providers to buyers at scale	Platform captures take rate + infrastructure margin; generates data advantage that compounds the commercial position

THE SUSTAINABLE TAKE RATE PRINCIPLE

The take rate must reflect value actually added — not the platform's ability to extract rent from captive participants

The history of marketplace monetization is full of platforms that raised take rates to the point where participants found ways to bypass them — Apple App Store's 30% rate triggered a decade of developer resentment and regulatory scrutiny; Amazon's seller fee increases have driven sellers to pursue direct-to-consumer alternatives. AI marketplaces face the same tension with a specific AI-economy twist: providers who are unhappy with the take rate can publish their models directly (the open-source ecosystem makes direct distribution viable for technical buyers) and buyers who feel over-charged can access models directly through foundation model APIs. The sustainable take rate is anchored to the value the marketplace genuinely adds — discovery, evaluation, trust, and commercial infrastructure — not to the marketplace's negotiating leverage.

Chapter Three — The Essentials

› Three participant types with different economic needs: providers (distribution + trust + infrastructure), buyers (discovery + evaluation + trust), platform (take rate + network data + incentive alignment).

› The take rate must reflect the infrastructure value the platform provides, not market power — AI marketplaces are more vulnerable to bypass than consumer marketplaces.

› Provider economics are most fragile: large providers have direct distribution alternatives; small providers have the highest dependency on marketplace distribution.

› The data network effect is the platform's most durable economic advantage: usage data from each transaction improves discovery, evaluation, and matching for all subsequent transactions.

› Incentive alignment is the foundation: a marketplace that grows at participants' expense will lose participants; a marketplace that grows with participants will compound.

PART TWO

Discovery Infrastructure

Taxonomy design. Evaluation architecture. The discovery engine.

CHAPTER FOUR

Building the Taxonomy: Making the Catalog Searchable at Scale

Task-based, function-based, and composition taxonomies. The ontology investment that makes discovery possible.

The taxonomy of B2B AI capabilities must be designed for the specific ways that B2B buyers think about their needs — which is different from how ML practitioners or academic researchers think about AI capabilities. The taxonomy design decision is the most consequential early architectural choice in B2B AI marketplace development.

B2B buyers in enterprise technology organizations think about AI capabilities in terms of business problems solved, workflows automated, and system integrations required — not in terms of model architectures, training methodologies, or academic benchmark

performance. A procurement team evaluating AI for accounts payable automation is asking: "Does this AI integrate with our SAP ERP? Does it handle the invoice formats our suppliers use? Does it meet our SOC 2 compliance requirements? What is the processing accuracy rate on our specific document types?" These questions require a taxonomy organized around business function, integration target, compliance certification, and workflow type — not around transformer architectures and benchmark datasets.

The enterprise application taxonomy organizes capabilities by the business problem they solve and the enterprise systems they integrate with. First dimension: business function (Finance, Legal, Human Resources, Customer Operations, Engineering, Procurement, Risk and Compliance, Sales and Marketing). Second dimension: workflow type within each function (Accounts Payable Automation, Contract Review, Employee Onboarding, Tier-1 Support Resolution, Code Review, Vendor Risk Assessment, Regulatory Filing, Lead Qualification). Third dimension: integration target (which enterprise systems does this AI connect with — SAP, Oracle, Workday, Salesforce, ServiceNow, Microsoft 365, or custom APIs). Fourth dimension: compliance posture (SOC 2, HIPAA, FedRAMP, ISO 27001, GDPR-ready, EU AI Act conformant).

This four-dimensional taxonomy is what enterprise buyers actually use to filter. "Show me AI capabilities for accounts payable automation that integrate with SAP and are SOC 2 certified" is a valid enterprise procurement query that the taxonomy must support. "Show me encoder-decoder transformer models fine-tuned on financial documents" is a valid ML practitioner query that is meaningless to the enterprise CFO approving the AP automation purchase.

The retail B2B taxonomy serves a different but overlapping buyer: the developer team that is building an AI application and needs to find the right component capabilities. This buyer wants to search by input/output type (what data formats does this accept and produce), API interface compatibility (does this support OpenAI-compatible APIs, Anthropic tool use, MCP servers), performance characteristics (what is the latency and

throughput at the scale I need), and commercial model (is there a free tier, what are the usage pricing tiers).

The critical design principle: a B2B AI marketplace must support both taxonomies simultaneously, presenting each buyer with the view of the catalog that matches their mental model. The enterprise IT leader searching for "HIPAA-compliant clinical documentation AI that integrates with Epic EHR" and the developer searching for "medical NLP model with FHIR R4 output schema" are potentially searching for the same capability — the taxonomy must surface it to both.

The composition taxonomy — the machine-readable capability description that allows AI agents to discover and combine capabilities programmatically — is the third layer that B2B AI marketplaces must eventually support. When an AI orchestration agent is assembling a workflow for contract analysis, it needs to find: a document ingestion capability (accepts PDF, outputs structured text), a legal NLP capability (accepts structured text, outputs issue annotations in JSON), a risk scoring capability (accepts issue annotations, outputs risk scores), and a report generation capability (accepts risk scores and original document, outputs a review memo in the required format). Each of these is a B2B marketplace query — programmatic, compositional, and commercially transactable.

AI Capability Taxonomy — Three Structures

Taxonomy type	Primary dimension	Secondary dimensions	Best suited for	Hugging Face implementation
Task-based	What does the AI do? (classification, generation, translation, detection)	Domain (medical, legal, financial) · Architecture (encoder, decoder, diffusion) · Modality (text, image, audio, multimodal)	Model hub archetype; developer buyers who think in terms of ML tasks	Tasks taxonomy with 50+ task types × domain × architecture tags — the most used filter on the Hub
Function-based	What business problem does this	Industry (healthcare, legal,	Enterprise application	AppExchange category tree:

	solve? (contract review, customer service, code generation)	finance) · Integration target (Salesforce, ServiceNow, SAP) · Complexity (simple, workflow, multi-agent)	marketplace archetype; business buyers who think in terms of business outcomes	Sales → AI Features → Lead Scoring; mapped to Salesforce platform objects
Composition-based	How does this capability compose with other capabilities? (inputs, outputs, schemas, protocols)	Interoperability (MCP-compatible, OpenAI-compatible, REST, GraphQL) · Data format (JSON, Parquet, PDF) · Compliance (HIPAA, FedRAMP, SOC2)	Agent-to-agent marketplace; AI orchestration frameworks; enterprise AI architects	Emerging — MCP tool manifest is the closest current standard; machine-readable capability cards not yet widespread

THE ONTOLOGY INVESTMENT

The taxonomy is an investment, not a feature — it requires active maintenance as AI capabilities evolve

The most common marketplace taxonomy mistake: design a taxonomy for the current state of AI capabilities and treat it as done. AI capabilities are evolving faster than any static taxonomy can accommodate. When multimodal models emerged, they did not fit cleanly into the text-only or image-only task categories. When AI agents emerged, they did not fit the single-model categories. The marketplace that treats taxonomy as a continuous investment — with a dedicated ontology team, community feedback mechanisms, and quarterly review cycles — will maintain discovery relevance as AI capabilities evolve. The marketplace that treats taxonomy as a launch-day deliverable will become unsearchable within two years.

Chapter Four — The Essentials

› Three taxonomy structures for different marketplace archetypes: task-based (ML practitioners), function-based (enterprise buyers), composition-based (AI agents and architects).

› The taxonomy determines whether the marketplace is discoverable as it scales — a poor taxonomy becomes the primary usability barrier at 10,000+ listings.

› Building a robust taxonomy requires ontology development, classification tooling for providers, and continuous refinement — not a one-time launch deliverable.

› Hugging Face's task-based taxonomy with 50+ task types is the most mature current example;

the composition-based taxonomy required for A2A commerce is still emerging.

› Machine-readable capability manifests — structured descriptions that AI agents can parse programmatically — are the evolutionary endpoint of the taxonomy investment.

CHAPTER FIVE

Evaluation Infrastructure: From Self-Reported to Independently Verified

Model cards. Platform benchmarks. Domain-specific attestation. Continuous runtime monitoring.

B2B AI capability evaluation has a fundamentally different standard from the developer community's evaluation model. Enterprise procurement teams are not satisfied with academic benchmark scores and community reviews. They require evidence that is specific to their deployment context, legally defensible in their compliance framework, and accountable to a named party — not to a leaderboard maintainer or a community reviewer.

The B2B evaluation stack has five layers organized by enterprise procurement requirement:

Self-reported technical documentation is the entry requirement, not the evaluation standard. Every listed capability must have documentation that describes its inputs, outputs, performance characteristics, training data provenance, intended uses, and known limitations. This documentation is the prerequisite for evaluation, not the evidence base for a procurement decision. An enterprise procurement team cannot purchase an AI capability without this documentation; they will not purchase solely on the basis of it.

Independent security certification is the first enterprise procurement gate. SOC 2 Type II is the baseline expectation for any B2B AI capability handling enterprise data. ISO

27001 is increasingly required for European enterprise buyers. FedRAMP is required for US government and regulated defense contractor deployments. Healthcare AI requires HIPAA technical safeguards and BAA capability. Financial services AI may require ISO 27001 and sector-specific security certifications. The B2B AI marketplace that requires and verifies these certifications as listing prerequisites eliminates the most time-consuming component of enterprise procurement: the individual vendor security review that IT teams must otherwise conduct for each new vendor.

Domain-specific performance attestation is the second procurement gate for enterprise buyers in regulated industries. Academic benchmarks measure general capability. Enterprise procurement teams need evidence specific to their use case. For legal AI, this means performance on the specific legal document types and jurisdictions the buyer operates in. For healthcare AI, it means performance on the clinical specialty and documentation format the buyer uses. For financial AI, it means performance on the specific financial instruments and reporting standards the buyer operates under. The B2B marketplace that develops and operates domain-specific attestation benchmarks — in partnership with industry organizations, academic institutions, and sector specialists — creates a trust infrastructure that no generic marketplace can quickly replicate.

Procurement document package is the commercial evaluation layer that enterprise buyers require before contract execution: a pre-completed vendor security questionnaire (VSQ), a data processing agreement (DPA) template compliant with GDPR and CCPA, liability and indemnification terms pre-negotiated to market standards, and a standard SLA covering uptime, performance, and data handling. Most B2B AI vendors cannot produce these documents quickly for every enterprise prospect — the marketplace that maintains standard procurement document packages for listed capabilities dramatically reduces the enterprise procurement cycle.

Runtime performance monitoring closes the evaluation loop after deployment: the marketplace monitors the deployed capability's performance against its documented standards and alerts the enterprise buyer when performance degrades. For B2B

enterprise buyers who have deployed AI in production systems, the marketplace's ongoing performance monitoring is as valuable as the pre-deployment evaluation — it converts the marketplace from a one-time purchase enabler into an ongoing quality assurance infrastructure for the enterprise's AI capability stack.

Evaluation Infrastructure — Four Layers					
Layer	**Trust level**	**Cost to produce**	**Operated by**	**Best example**	**Buyer action**
Self-reported metrics	Low — provider claims without independent verification	Low (documentation time)	Provider	Hugging Face model cards — structured documentation of capabilities, training data, evaluation results, limitations	Informational only; verify independently before production deployment
Platform-verified benchmarks	Medium — standardized benchmarks run by neutral party	Medium (compute for benchmark runs at scale)	Marketplace platform	Hugging Face Open LLM Leaderboard — standardized benchmark suite run against all submitted models; results published by Hugging Face	Use for initial shortlisting and capability comparison; still insufficient for enterprise production deployment without further evaluation
Domain-specific performance attestation	High — independent third-party evaluation against domain-authoritative benchmarks	High (requires domain expertise and authoritative benchmark data)	Third-party evaluation organizations partnering with marketplace	Clinical NLP benchmarks by academic medical centers; legal reasoning benchmarks by law school AI programs; financial calculation benchmarks by accounting firms	Sufficient for most enterprise production deployment decisions; required for regulated industry deployment

Continuous runtime monitoring	Highest — real-world performance validation in the buyer's actual deployment context	Ongoing operational cost; requires buyer system integration	Marketplace platform + buyer (co-operated)	Replicate's latency and throughput monitoring; emerging: quality-in-production monitoring for production AI deployments	Provides ongoing confidence that deployed model is performing as initially evaluated; alerts on performance degradation

CASE STUDY: HUGGING FACE OPEN LLM LEADERBOARD *Platform-Verified Benchmarking at Ecosystem Scale*	
What it is	An independent, community-maintained benchmark that evaluates language models on standardized academic benchmarks including HellaSwag, ARC, TruthfulQA, MMLU, and others — run by Hugging Face against all submitted models with transparent methodology and results published publicly.
Commercial impact	When Mistral 7B appeared on the leaderboard in September 2023 outperforming models ten times larger, it generated 1M+ downloads within weeks and established Mistral AI as a serious foundation model competitor — purely through leaderboard performance, without traditional enterprise sales or marketing investment. The leaderboard created the discovery signal that turned technical performance into commercial momentum.
The trust mechanism	The leaderboard's value is precisely that it is not operated by the model providers being evaluated. An OpenAI GPT-4o result on OpenAI's own benchmark is less credible than the same model's result on Hugging Face's independently run benchmark. The marketplace operator's neutrality is what makes the evaluation infrastructure commercially valuable.
The limitation	Academic benchmarks do not measure production performance for specific enterprise use cases. A model that tops the MMLU benchmark may underperform a smaller, domain-specific model on clinical documentation tasks. The Open LLM Leaderboard is necessary but not sufficient for enterprise deployment decisions — the domain-specific attestation layer is still required.
The evolution	Hugging Face's 2024 launch of the Open Medical LLM Leaderboard, the Open Legal Benchmark, and the Open Code LLM Leaderboard represents the extension of the platform-verified benchmarking approach into domain-specific evaluation — the layer required for enterprise buyers in regulated industries.

Chapter Five — The Essentials

› Four evaluation layers: self-reported (low trust), platform-verified (medium trust), domain-specific attestation (high trust), continuous runtime monitoring (highest trust).

› The leaderboard discovery effect: independent platform benchmarks can generate more commercial momentum for a new model than any amount of traditional marketing.

› Domain-specific attestation is required for enterprise deployment in regulated industries — healthcare, legal, financial, and government buyers require more than generic academic benchmarks.

› Continuous runtime monitoring is the emerging capability that converts the marketplace from a one-time purchase enabler to an ongoing quality assurance infrastructure.

› The trust infrastructure is the competitive moat — a marketplace with stronger evaluation infrastructure attracts better-quality providers (they have confidence their quality will be recognized) and higher-trust buyers (enterprises that need production confidence).

CHAPTER SIX

B2B Discovery Channels: Self-Serve, Enterprise-Assisted, Partner-Led, Agent-Driven

Three discovery modes. The recommendation system. Compositional search for AI agents.

B2B AI capability discovery happens through four distinct channels, each requiring different marketplace infrastructure and serving different buyer behaviors. The B2B AI marketplace must support all four simultaneously.

The self-serve discovery channel serves the retail B2B buyer: an engineer, data scientist, or product manager at a business organization who discovers the marketplace through web search, community referral, or technology publication, evaluates capabilities through the marketplace's online interface, and makes a purchasing decision independently within their delegated authority. The marketplace's investment for this channel is in SEO and content (making capabilities discoverable through organic search), UI/UX (making the evaluation experience productive for a technically

sophisticated B2B buyer who is self-directing the evaluation), and retail commercial infrastructure (sign-up flow, credit card or purchase order billing, standard terms acceptance).

The enterprise-assisted discovery channel serves the enterprise B2B buyer: an IT organization, a procurement team, or an executive-sponsored initiative that has identified an AI capability requirement and is conducting a formal evaluation. The marketplace's investment for this channel is in procurement support materials (vendor profiles, compliance documentation, security questionnaire responses), enterprise sales liaison (a marketplace representative who can facilitate introductions between enterprise buyers and listed vendors), and the Private Offer infrastructure that allows custom commercial terms to be executed through the marketplace's billing infrastructure.

The partner-led discovery channel serves B2B buyers who discover marketplace capabilities through system integrators, technology consultants, and managed service providers who are evaluating AI capabilities on their clients' behalf. This channel is particularly important in enterprise B2B markets because a significant fraction of enterprise AI deployment decisions involve a consulting partner who recommends and often implements the AI capability. The marketplace must support the partner channel with partner portal tools (deal registration, commission tracking, client portfolio management), partner certification programs, and commercial structures that allow partners to bundle marketplace capabilities into managed service offerings.

The agent-driven discovery channel is the emerging channel where AI orchestration agents query the marketplace programmatically to identify and evaluate capabilities for automated workflows. This channel serves the agent-buyer customer segment described in the previous chapter. The marketplace's investment for this channel is in the machine-readable API infrastructure: capability manifests, real-time pricing endpoints, performance attestation APIs, and micropayment settlement — all of which must be available through programmatic interfaces, not just through the marketplace's human-facing web interface.

Three Discovery Modes — Design Requirements				
Mode	**Buyer behavior**	**Technical requirement**	**Maturity**	**Commercial value**
Intent-based search	Buyer knows what they need; queries for it in natural language or structured filters	Semantic search that bridges vocabulary gap between buyer language (business terms) and provider language (ML terms); taxonomy-guided filtering	Mature — Hugging Face, AWS Marketplace, AppExchange all have functional intent-based search	Medium — solves the known-need discovery problem; does not surface what the buyer did not know to look for
Context-aware recommendation	Buyer has been using capabilities; marketplace surfaces better alternatives or complementary capabilities proactively	User behavior modeling + performance comparison + new listing alerts calibrated to individual buyer's usage context	Early — Hugging Face trending models is basic version; personalized recommendation based on specific performance requirements is not yet widespread	High — creates discovery value beyond what the buyer would generate through active search; drives organic capability expansion
Compositional discovery	Buyer (or their agent) has a complex multi-step task requirement; marketplace identifies the combination of capabilities that can satisfy it	Capability manifest standard + capability graph (which capabilities can be composed with which) + query expansion that decomposes complex requirements into capability components	Emerging — not yet available in current commercial AI marketplaces; the MCP tool ecosystem is the closest analog	Very high — enables AI agents to discover and assemble capability workflows dynamically; the primary discovery mode for A2A marketplace commerce

Chapter Six — The Essentials

› Three discovery modes serve different buyer behaviors: intent-based (known need), context-aware (unknown opportunity), compositional (complex workflow).

› Hugging Face Spaces convert technical model listings into accessible demonstrations — the most effective discovery pathway for non-technical enterprise buyers.

› Compositional discovery is the mode that enables A2A marketplace commerce: the capability graph and machine-readable manifests are the investment required.

› Context-aware recommendation creates discovery value that compounds with platform usage data — each additional transaction improves recommendation quality for all buyers.

› The vocabulary gap between buyer language and provider language is the primary friction point in intent-based search — semantic understanding is not optional for discovery at scale.

PART THREE

Pricing Models for AI Marketplaces

Seven provider pricing models. Take rate architecture. Dynamic pricing for agent commerce.

CHAPTER SEVEN

Provider Pricing Models: Seven Structures for AI Capabilities

Per-inference to enterprise tier. The open-claw challenge at each layer. Composition pricing.

B2B AI capability pricing must serve two commercially distinct buyers — the retail B2B buyer who self-selects from catalog pricing, and the enterprise B2B buyer who negotiates custom commercial terms — while maintaining the pricing consistency and transparency that makes the marketplace trustworthy to both.

The seven pricing models applicable to B2B AI marketplaces differ from consumer marketplace pricing in three critical ways: the minimum transaction size is larger (B2B contracts are measured in thousands to millions of dollars annually, not dollars monthly), the procurement process is slower (enterprise B2B procurement takes weeks

to months, not seconds), and the commercial relationship is more complex (MSAs, DPAs, SLAs, and compliance certifications are standard requirements, not optional add-ons).

Per-API-call / per-inference pricing serves retail B2B buyers who want variable-cost access to AI capabilities without usage commitment. In a B2B context, this pricing is typically offered through a metered billing arrangement attached to the buyer's organizational account — not through individual developer accounts as in consumer API marketplaces. The B2B billing context means that the cost aggregates to a line item in the organization's technology spend report, is attributable to a specific engineering team or project, and is subject to the organization's spend governance controls. The commercial contract governing per-API-call pricing in B2B is typically a terms of service plus a DPA — lighter than an enterprise agreement, but still a commercial contract between businesses.

The retail B2B subscription tier (monthly or annual commitment with defined usage allocation) is the primary commercial structure for retail B2B customers at most AI marketplaces. The tier includes: a defined monthly or annual fee (predictable for the buyer's technology budget), a defined usage allocation (number of API calls, inference requests, or processing volume), standard commercial terms (ToS + DPA), SLA covering availability and support response time, and data handling commitments appropriate for business-to-business data processing. Overages above the allocation are billed at a defined per-unit rate. The subscription tier is the commercial structure that allows retail B2B buyers to make a commercial commitment with organizational accounting without requiring full enterprise procurement processes.

The enterprise B2B annual contract structure is the commercial model for the large-organization segment. The enterprise contract includes: negotiated pricing with volume discounts, custom SLA with defined penalties for SLA breach, bespoke DPA terms appropriate for the enterprise's specific regulatory environment, BAA (business associate agreement) for healthcare buyers, government contract addenda for public sector buyers, liability caps and indemnification terms negotiated to the enterprise's

standards, and dedicated support entitlements. Enterprise contracts are executed through the marketplace's Private Offer infrastructure (on cloud provider marketplaces) or through the marketplace operator's enterprise sales motion with billing through the marketplace's standard invoice system.

The capacity commitment model — where the enterprise B2B buyer commits to a defined volume of capability consumption over a defined period in exchange for a unit price discount — is particularly important for B2B AI marketplace buyers whose AI deployment is scaling rapidly and who want budget certainty. The capacity commitment is the B2B equivalent of the reserved instance model in cloud computing: the buyer commits to consuming a defined volume, the vendor commits to providing that capacity, and both parties benefit from the predictability. Overages above the committed capacity are available at the standard per-unit rate without requiring a new contract.

The platform integration fee is a B2B-specific pricing model that exists in enterprise application marketplaces: the buyer pays an integration fee that covers the professional services required to connect the AI capability to their existing enterprise systems, in addition to the ongoing subscription or usage fees. In B2B AI deployments, the integration work — connecting the AI to the buyer's data sources, configuring it for the buyer's specific document formats and workflows, testing it against the buyer's production data — often requires more commercial value than the raw capability access. The platform integration fee structure recognizes this by separating the integration value from the ongoing capability access value.

The outcome-based commercial model — where the B2B buyer pays based on verified business outcomes rather than capability access or usage — is the highest-value and least commoditizable commercial structure for B2B AI marketplaces. A legal AI service that charges per contract review delivered (regardless of token consumption) or a financial AI service that charges per accurately categorized transaction (regardless of processing cost) is anchored to the value the capability creates for the B2B buyer's business, not to the cost of delivering it. The open-claw effect (Book 8) is minimized

under outcome-based pricing because AI capability improvement increases outcome reliability and volume rather than eroding per-unit revenue.

Seven AI Marketplace Pricing Models				
Model	**Billing unit**	**Suited for**	**Open-claw risk**	**Representative platform**
Per-inference	Token / second / image / request	API-accessible AI capabilities with variable usage; developer buyers	High — efficiency improvement reduces cost per inference while value per inference grows; revenue declines as AI improves	Hugging Face Inference API, Replicate, fal.ai, OpenAI API
Subscription with inference allocation	Monthly fee + overage above allocation	Enterprise buyers needing budget predictability; moderate to high usage volumes	Low — subscription provides revenue floor; overage captures growth above committed level	Hugging Face Enterprise Hub, Anthropic Claude Teams, AWS Marketplace SaaS subscriptions
Fine-tuning and training	Per-GPU-hour for training compute + per-inference for resulting model	Enterprises needing domain-specific model adaptation; providers offering specialization services	Medium — compute efficiency improvement reduces training cost; fine-tune value is customer-specific and less subject to commodity pressure	Replicate fine-tuning, AWS SageMaker, Hugging Face AutoTrain
Dataset licensing	Per-download / subscription	Data providers with	Low — dataset value is in provenance	Hugging Face Datasets commercial

	/ usage-based	proprietary training or evaluation datasets; enterprises needing specific data assets	and quality, not inference efficiency	licenses, Scale AI dataset marketplace, Roboflow
Capability composition	Per-workflow / per-outcome for multi-capability service	Enterprise buyers deploying complex AI workflows; marketplace-assembled service packages	Very low — composition value exceeds sum of components; outcome pricing applies when composition delivers verifiable outcomes	Early stage — not yet widespread; Salesforce Flow AI, ServiceNow AI workflows are early examples
Enterprise tier	Annual contract with SLA, compliance, support	Regulated industry buyers; large enterprises requiring governance, security, auditability	Low — enterprise tier value is in accountability and compliance infrastructure, not AI performance alone	Hugging Face Enterprise, AWS Marketplace Private Offers, Salesforce AppExchange enterprise tiers
Freemium to commercial conversion	Free (community tier) + paid (commercial tier above threshold)	Open-source models targeting developer adoption before commercial conversion	Strategic — free tier builds network effects that sustain commercial tier value; take rate on community is zero	Meta Llama commercial licenses, Mistral AI commercial licenses, Stability AI enterprise

THE EFFICIENCY TRAP FOR PER-INFERENCE PRICING

As models become more efficient, per-inference pricing systematically underprices value — the open-claw problem at the model layer

A model that generates the same quality output in 500 tokens that previously required 2,000 tokens has improved 4× in efficiency. Under per-token pricing, the provider's revenue for that output falls by 75% while the buyer's value received is unchanged. This is the token-layer open claw described in Book 8. The closing strategy for per-inference providers: shift pricing toward output quality metrics (did the output meet the defined quality standard?) rather than input consumption metrics (how many tokens were consumed to produce it?). Replicate's quality-augmented pricing experimentation and Anthropic's capability-tiered API pricing are early attempts at this shift.

Chapter Seven — The Essentials

› Seven pricing models; each has different open-claw risk profiles — per-inference has the highest, enterprise tier and capability composition have the lowest.

› The fundamental principle: anchor price to value delivered (outcomes, capabilities, quality), not to cost of delivery (tokens, seconds, compute).

› Freemium is a strategic investment in network effects, not a pricing failure — the take rate on community usage is zero, but the network effect it builds compounds into commercial tier value.

› Capability composition pricing is the highest-value, least commoditizable model: value the outcome the combination delivers, not the sum of the component costs.

› Enterprise tier pricing is anchored to compliance and accountability value — it is largely immune to AI efficiency improvement because compliance infrastructure cost is not primarily driven by inference cost.

CHAPTER EIGHT

The Take Rate: Sustainable Monetization for Marketplace Operators

Setting the rate. Tiered structures. Value-added services that justify premiums. Bypass risk.

The B2B marketplace take rate — the percentage of transaction value the marketplace operator retains — has specific dynamics in B2B contexts that differ from consumer marketplace economics.

B2B marketplace take rates are generally lower than consumer marketplace take rates for the same category of software because: B2B buyers have more procurement power and will negotiate around high take rates by establishing direct vendor relationships; B2B transactions are larger individually (justifying more negotiation effort to avoid the take rate); and B2B vendors at scale have more viable direct distribution alternatives than consumer app developers (enterprise sales teams and direct procurement relationships are commercially feasible).

The sustainable B2B take rate range is 8–20% for standard transactions, with the specific rate dependent on the value-add the marketplace provides:

8–12% is appropriate when the marketplace primarily provides discovery and transaction infrastructure, with limited trust or compliance value-add. This rate reflects the value of reaching a concentrated B2B buyer audience that the vendor could not efficiently reach independently, without significant trust infrastructure.

12–18% is appropriate when the marketplace provides discovery, security certification, compliance documentation, standard procurement packages, and enterprise sales support. The additional services justify the premium by reducing the vendor's cost of enterprise sales significantly — a vendor that would otherwise spend six months on each enterprise procurement cycle gets that cycle reduced to weeks through the marketplace's pre-clearance infrastructure.

18–25% is defensible when the marketplace controls a captive B2B buyer audience (within a platform ecosystem like Salesforce or ServiceNow) and provides a security certification that is the de facto procurement prerequisite for that buyer segment. At this level, the marketplace is providing genuine commercial infrastructure that the vendor cannot replicate — the Salesforce AppExchange security badge is worth 18–25% of transaction value because it unlocks enterprise procurement approvals that would cost the vendor far more in direct sales effort.

Private Offer structures on cloud marketplaces effectively reduce the take rate for large enterprise transactions: the cloud provider charges a lower rate for high-value Private

Offer transactions (often 3–8%) because the provider's primary commercial interest is in the infrastructure consumption that enterprise AI deployments generate, not in the take rate on the AI service subscription itself.

The tiered take rate structure that aligns with B2B marketplace economics: zero for community/free-tier listings (the supply network investment), 12–15% for retail B2B standard subscriptions (where marketplace discovery and trust infrastructure provide genuine value), and 8–10% for enterprise B2B transactions above a defined threshold (where the per-transaction value is high enough that a lower percentage still generates significant absolute revenue, and where direct vendor relationships are more likely alternatives if the rate is too high).

Take Rate Benchmarks and Justification

Take rate	What it requires to be sustainable	Examples	Risk if unjustified
5–10%	Basic connectivity and transaction processing; minimal trust or discovery infrastructure	Basic API marketplaces; early-stage developer tool directories	Low bypass risk but low platform value — participants stay because switching is minor, not because the platform is valuable
10–15%	Good discovery infrastructure (search, taxonomy, recommendations); platform-managed transaction mechanics; basic community features	AWS Marketplace data category; developer tool marketplaces	Moderate bypass risk — providers with significant volume have incentive to pursue direct distribution; larger providers may negotiate lower rates
15–20%	Strong discovery + security/compliance certification + developer tools + enterprise sales support + SLA management infrastructure	Salesforce AppExchange (15–25%); Shopify App Store (20%)	Sustainable when trust infrastructure is genuinely valuable; providers who cannot replicate the trust infrastructure independently stay on the platform
20–30%	All of the above +	Apple App Store	High bypass risk if

	exclusive reach (platform-captive buyer base) + significant value-add services (compliance certification, enterprise sales, customer success)	(15–30%); Salesforce premium AppExchange tiers	platform lock-in weakens; sustainable only when platform controls distribution access (iOS app store model)

Tiered Take Rate Structure — Design Reference

Revenue tier	Take rate	Rationale	Applies to	Implementation
$0 – $50K annual revenue on platform	0%	Community-building tier; encourages early-stage providers; builds supply network	New providers establishing their marketplace presence; community/free models	Track revenue; automatic transition when threshold crossed
$50K – $500K annual revenue	15%	Base commercial tier; provider has demonstrated marketplace viability	Established commercial providers with meaningful transaction volume	Standard take rate tier
$500K – $5M annual revenue	12%	Volume discount reflecting relationship value and platform commitment	Providers generating meaningful marketplace revenue; candidates for co-marketing	Negotiated or automatic based on revenue tier
>$5M annual revenue	8–10%	Strategic partnership tier; provider is a significant platform participant	Top providers representing significant platform GMV; strategic ecosystem partners	Negotiated individually; may include co-marketing, dedicated support, certification priority

Chapter Eight — The Essentials

› Sustainable take rates: 10–15% for discovery-only value; 15–20% with compliance certification and enterprise infrastructure; 20–30% only with genuine platform-controlled distribution.

› Tiered take rates that decline with revenue growth are the standard for major platforms — they align platform incentives with provider success and reduce bypass incentives at scale.

› The bypass risk is higher for AI marketplaces than for consumer platforms: technical buyers can access models directly through APIs; commercial terms can be replicated without marketplace infrastructure.

› Value-added services (certification, compliance, enterprise sales support, SLA management) are the commercial justification for take rates above 15% — without them, the take rate is rent extraction.

› The key question at every pricing decision: are we charging for value we provide, or for friction we create? The first is sustainable; the second will be bypassed as participants grow sophisticated.

CHAPTER NINE

Dynamic Pricing and the Agent Commerce Protocol

Real-time pricing APIs. Machine-readable manifests. The Dynamic AI Capability Pricing Protocol.

The Model Context Protocol (MCP), developed by Anthropic and introduced in Book 4, is the technical standard that governs how AI agents interact with external tools and capabilities. The integration between MCP servers and B2B AI marketplaces creates the technical infrastructure for B2B agent commerce — with the B2B governance requirements described in the previous chapter implemented at the protocol layer.

The B2B MCP marketplace integration model works as follows. A B2B AI capability provider creates an MCP server that exposes their capabilities as structured tools — with defined input schemas, output schemas, and capability manifests. The B2B AI marketplace wraps this MCP server with a commercial middleware layer that enforces B2B governance requirements: organizational account validation (the calling agent is registered to a valid organizational account), spending authority verification (the transaction does not exceed the agent's configured authority), vendor relationship verification (the organizational account has a valid commercial relationship with this provider), data handling enforcement (the submitted data complies with the DPA

between the organization and this provider), and transaction logging (the transaction is recorded in the organizational account's audit log with full attribution).

The MCP tool registry — the B2B marketplace's catalog of available MCP-exposed AI capabilities — is the discovery layer for agent-driven B2B commerce. An AI orchestration agent that needs to assemble a multi-step workflow queries the MCP tool registry to identify available capabilities by capability type, interface compatibility, performance characteristics, and commercial availability. The registry returns a structured list of candidate capabilities with their manifests, pricing, and performance data — allowing the agent to select the optimal capability for its current task programmatically.

The commercial layer that B2B marketplace operators must add to MCP integration goes beyond what the open-source MCP protocol specifies. MCP defines the technical interaction between agents and tools — the message formats, the error handling, the streaming patterns. It does not define the commercial layer — the pricing, the settlement, the authority governance, or the compliance enforcement. B2B marketplace operators must build this commercial layer as a wrapper around the technical MCP protocol, making B2B commercial requirements transparent to the AI agent's task logic while enforcing them at the infrastructure layer.

The Hugging Face MCP integration launched in 2025 — which makes every Hugging Face Space available as an MCP server — represents the first significant deployment of marketplace-connected MCP at ecosystem scale. In its current form, the commercial layer is missing: MCP invocations of Hugging Face Spaces are free or covered by existing subscriptions, with no per-invocation B2B commercial settlement. The next phase — adding B2B commercial settlement, organizational account enforcement, and spending authority governance to the MCP integration layer — is the infrastructure investment that converts Hugging Face MCP from a developer tool into a B2B agent commerce platform.

Dynamic AI Capability Pricing Protocol — Required Components

Component	Function	Technical specification	Current maturity	Commercial importance
Capability manifest	Machine-readable description of the capability: inputs, outputs, performance characteristics, usage constraints	JSON-LD structured data: capability type, input schema, output schema, performance metrics, license terms, pricing model type	Early — Hugging Face model cards are semi-structured; MCP tool manifests are structured but not commercially annotated	Critical — without machine-readable manifests, agents cannot evaluate capabilities programmatically
Real-time pricing endpoint	API returning current price for a specific invocation type at the moment of query	REST endpoint: POST /capabilities/{id}/price with invocation parameters; returns: price (amount, currency), validity_window (seconds), terms_url	Emerging — Replicate provides per-model pricing via API; dynamic real-time pricing not yet standard	High — enables agents to make price-aware purchasing decisions; required for cost-constrained agent workflows
Performance attestation endpoint	API returning current verified performance metrics for the capability	REST endpoint: GET /capabilities/{id}/performance with optional context parameters; returns: benchmark scores, latency p50/p95, error rate, last_verified timestamp	Early — Replicate provides latency metrics; quality metrics in real-time are not yet standard	High — enables agents to make quality-aware capability selection; prevents degraded capabilities from being auto-selected
Transaction settlement	Execute and confirm the capability transaction, deducting from the buyer's account and logging the invocation	REST endpoint: POST /transactions with: capability_id, invocation_params, buyer_agent_id, max_price; returns: transaction_id, actual_price, invocation_token	Partial — prepaid credit accounts enable instant settlement; per-invocation settlement not standard	Critical — A2A marketplace commerce requires instant settlement; invoice-based settlement is incompatible with machine-speed transactions

Audit log	Immutable record of every agent invocation for billing verification, compliance, and dispute resolution	Append-only log: timestamp, agent_id, capability_id, invocation_params_hash, response_hash, price, transaction_id	Good — most cloud API providers log invocations; A2A-specific audit trails not yet standard	High — required for enterprise governance of agent spending; enables dispute resolution when agent-invoked capabilities produce unexpected results

Chapter Nine — The Essentials

› The Dynamic AI Capability Pricing Protocol has five components: manifest, real-time pricing, performance attestation, transaction settlement, and audit log.

› Current maturity is partial — each component exists in some form, but a unified standard enabling seamless A2A marketplace commerce does not yet exist.

› The marketplace that builds this complete protocol stack first creates the infrastructure that AI orchestration frameworks will standardize around.

› Machine-readable manifests are the prerequisite for everything else — without them, agents cannot evaluate capabilities programmatically, making all other protocol components inaccessible.

› Instant settlement (prepaid credit accounts or streaming micropayments) is required for A2A commerce — invoice-based settlement operates at human speed, not machine speed.

PART FOUR

Trust, Compliance, and IP Governance

The trust stack. Regulatory compliance. IP in an AI marketplace.

CHAPTER TEN

Trust Architecture: Five Layers of AI Marketplace Governance

Model cards to runtime monitoring. Safety evaluation. Dispute resolution.

B2B AI marketplace trust infrastructure must meet a higher standard than developer community trust infrastructure because the consequences of deploying an unsafe, non-compliant, or underperforming AI capability in an enterprise production environment can be severe, legally significant, and organizationally disruptive.

The B2B enterprise buyer's trust requirements are governed by four organizational functions, each with distinct standards:

The IT security team requires: SOC 2 Type II certification (or equivalent), vendor security questionnaire responses, penetration testing documentation, data breach notification procedures, and demonstrated compliance with the organization's information security policy. The IT security team's primary concern is data handling — where is the organization's data processed, who has access to it, what happens in a breach, and can the vendor be removed cleanly from the organization's data environment if the relationship ends.

The legal and compliance team requires: DPA with the specific regulatory language applicable to the organization's jurisdiction and industry, liability terms that are appropriate for the potential consequences of AI errors, IP representations covering training data provenance and output ownership, and compliance certifications specific to the industry (BAA for healthcare, regulatory compliance for financial services, FedRAMP for government). The legal team's primary concern is contractual liability and regulatory compliance — can the organization deploy this AI without creating legal exposure?

The procurement team requires: formal vendor registration in the organization's vendor management system, standard commercial terms (pricing schedule, payment terms, volume discounts), and a commercial relationship that fits the organization's

procurement processes (PO-based billing, enterprise agreement, subscription invoice). The procurement team's primary concern is commercial efficiency and governance — does this vendor relationship fit our procurement system, and is the pricing structure auditable and governable?

The business unit deploying the AI requires: evidence that the AI performs well for their specific use case (not just general benchmarks), integration compatibility with their existing systems, clear escalation paths for issues, and training support to build the internal capability to manage the AI deployment. The business unit's primary concern is operational success — will this AI actually work for our specific workflows, and do we have the support required to make it successful?

The B2B AI marketplace's trust infrastructure must address all four sets of requirements simultaneously. This is why B2B trust infrastructure is more complex and more expensive to build than developer community trust infrastructure — the community developer needs to know if the model works; the enterprise B2B buyer needs to know if it is safe, legal, governable, and operational.

The trust infrastructure layers for B2B AI marketplaces:

Pre-listing certification covers IT security and legal/compliance requirements: security certification verification, DPA template provision, liability terms review, training data provenance documentation, and compliance certification verification. Every listing must pass pre-listing certification requirements before being visible to enterprise B2B buyers. The marketplace operates a vendor certification program that verifies these requirements and maintains the certification status, alerting enterprise buyers when certifications expire or when a listed capability's compliance posture changes.

Procurement documentation package covers procurement team requirements: a pre-completed vendor security questionnaire, standard commercial terms at each pricing tier, volume discount schedules, and payment term documentation. The marketplace maintains and updates these documents for each listed capability, eliminating the

vendor-by-vendor documentation collection that enterprise procurement teams otherwise perform for each new vendor.

Domain performance attestation covers business unit requirements: independent evaluation of capability performance on domain-specific benchmarks that are authoritative in the buyer's industry. The B2B marketplace operates or commissions domain-specific benchmarks in partnership with industry organizations — clinical benchmarks for healthcare AI, legal reasoning benchmarks for legal AI, financial calculation benchmarks for financial AI, supply chain optimization benchmarks for industrial AI. Performance results are published with the evaluation methodology, the reference dataset, and the evaluation partner's credentials.

Runtime performance monitoring covers ongoing operational requirements: continuous monitoring of deployed capability performance against the standards documented at listing, with automated alerts to the enterprise buyer when performance degrades. This monitoring converts the marketplace from a one-time procurement tool to an ongoing operational quality assurance infrastructure — which is the commercial relationship that generates the highest retention and the highest NRR.

AI Marketplace Trust Architecture — Five Layers					
Layer	**What it provides**	**How it is operated**	**Trust level**	**Investment required**	**Enterprise buyer impact**
1. Model card standards	Documentation of capability, limitations, training data, evaluation methodology	Provider creates; marketplace verifies accuracy for certified listings	Low-medium baseline	Low (documentation tools + verification spot-checks)	Necessary but not sufficient; provides information for initial evaluation
2. Platform-verified benchmarks	Independent performance measurement on standardized benchmarks by neutral party	Marketplace runs benchmark suite; publishes results independently	Medium — controlled by neutral party; results not cherry-picked	Medium (compute for benchmark runs at scale + benchmark curation)	Enables systematic comparison; creates the leaderboard discovery effect

		of provider claims			
3. Domain-specific attestation	Expert evaluation against authoritative domain benchmarks	Specialist organizations (medical schools, law programs, security firms) partner with marketplace to operate domain benchmarks	High — domain expert authority + marketplace neutrality	High (partnership development + benchmark dataset curation + ongoing operations)	Required for regulated industry production deployment; enables enterprise procurement approval
4. Runtime monitoring	Continuous performance validation in production; drift detection; degradation alerting	Marketplace provides monitoring agent; buyer integrates with production deployment; alerts flow to both parties	Highest — real-world performance in actual production context	Ongoing operational cost + buyer integration requirement	Converts marketplace from purchase enabler to ongoing quality assurance infrastructure; highest-value enterprise feature
5. Dispute resolution	Commercial backstop when trust failures occur despite preceding layers	Marketplace mediates disputes; provides audit trail as evidence; applies remedies (refund, credit, listing removal)	Procedural — quality of the mechanism matters as much as the mechanism's existence	Legal framework + dispute mediation process + remedies policy	Required for enterprise procurement approval in most organizations; enables commercial accountability for AI capability failures

Chapter Ten — The Essentials

› Five trust layers: model cards (documentation), platform benchmarks (independent verification), domain attestation (expert evaluation), runtime monitoring (production validation), dispute resolution (commercial backstop).

› Enterprise buyers require at minimum layers 1–3 for initial production deployment; layers 4–5 are required for ongoing production reliance and regulated industry deployment.

› The trust infrastructure is the competitive moat: it takes years to build to enterprise grade, cannot be purchased, and compounds in value as more listings are certified and more

deployment data accumulates.

› Domain-specific attestation (layer 3) is the most underinvested layer currently — the biggest opportunity for new marketplaces to differentiate in specific enterprise verticals.

› The marketplace that implements all five layers for a specific vertical before competitors builds an enterprise trust moat that makes displacement extremely difficult.

CHAPTER ELEVEN

Compliance, Regulation, and IP Governance

EU AI Act. Healthcare and sector compliance. Model output ownership. Training data provenance.

B2B AI marketplace trust infrastructure must meet a higher standard than developer community trust infrastructure because the consequences of deploying an unsafe, non-compliant, or underperforming AI capability in an enterprise production environment can be severe, legally significant, and organizationally disruptive.

The B2B enterprise buyer's trust requirements are governed by four organizational functions, each with distinct standards:

The IT security team requires: SOC 2 Type II certification (or equivalent), vendor security questionnaire responses, penetration testing documentation, data breach notification procedures, and demonstrated compliance with the organization's information security policy. The IT security team's primary concern is data handling — where is the organization's data processed, who has access to it, what happens in a breach, and can the vendor be removed cleanly from the organization's data environment if the relationship ends.

The legal and compliance team requires: DPA with the specific regulatory language applicable to the organization's jurisdiction and industry, liability terms that are

appropriate for the potential consequences of AI errors, IP representations covering training data provenance and output ownership, and compliance certifications specific to the industry (BAA for healthcare, regulatory compliance for financial services, FedRAMP for government). The legal team's primary concern is contractual liability and regulatory compliance — can the organization deploy this AI without creating legal exposure?

The procurement team requires: formal vendor registration in the organization's vendor management system, standard commercial terms (pricing schedule, payment terms, volume discounts), and a commercial relationship that fits the organization's procurement processes (PO-based billing, enterprise agreement, subscription invoice). The procurement team's primary concern is commercial efficiency and governance — does this vendor relationship fit our procurement system, and is the pricing structure auditable and governable?

The business unit deploying the AI requires: evidence that the AI performs well for their specific use case (not just general benchmarks), integration compatibility with their existing systems, clear escalation paths for issues, and training support to build the internal capability to manage the AI deployment. The business unit's primary concern is operational success — will this AI actually work for our specific workflows, and do we have the support required to make it successful?

The B2B AI marketplace's trust infrastructure must address all four sets of requirements simultaneously. This is why B2B trust infrastructure is more complex and more expensive to build than developer community trust infrastructure — the community developer needs to know if the model works; the enterprise B2B buyer needs to know if it is safe, legal, governable, and operational.

The trust infrastructure layers for B2B AI marketplaces:

Pre-listing certification covers IT security and legal/compliance requirements: security certification verification, DPA template provision, liability terms review, training data provenance documentation, and compliance certification verification. Every listing

must pass pre-listing certification requirements before being visible to enterprise B2B buyers. The marketplace operates a vendor certification program that verifies these requirements and maintains the certification status, alerting enterprise buyers when certifications expire or when a listed capability's compliance posture changes.

Procurement documentation package covers procurement team requirements: a pre-completed vendor security questionnaire, standard commercial terms at each pricing tier, volume discount schedules, and payment term documentation. The marketplace maintains and updates these documents for each listed capability, eliminating the vendor-by-vendor documentation collection that enterprise procurement teams otherwise perform for each new vendor.

Domain performance attestation covers business unit requirements: independent evaluation of capability performance on domain-specific benchmarks that are authoritative in the buyer's industry. The B2B marketplace operates or commissions domain-specific benchmarks in partnership with industry organizations — clinical benchmarks for healthcare AI, legal reasoning benchmarks for legal AI, financial calculation benchmarks for financial AI, supply chain optimization benchmarks for industrial AI. Performance results are published with the evaluation methodology, the reference dataset, and the evaluation partner's credentials.

Runtime performance monitoring covers ongoing operational requirements: continuous monitoring of deployed capability performance against the standards documented at listing, with automated alerts to the enterprise buyer when performance degrades. This monitoring converts the marketplace from a one-time procurement tool to an ongoing operational quality assurance infrastructure — which is the commercial relationship that generates the highest retention and the highest NRR.

Compliance Framework for AI Marketplace Providers

Framework	Jurisdiction	Who it covers	Marketplace requirement	Provider requirement	Buyer benefit
EU AI Act	European Union	AI systems operating in	Marketplace must not list	High-risk AI systems	Confidence that listed

		the EU affecting EU persons	prohibited systems; must document how it verifies high-risk system conformity assessments	require conformity assessment before listing; documentation for regulatory inspection	capabilities comply with EU legal requirements; reduced regulatory risk for EU deployment
FDA SaMD Framework (510k/De Novo/PMA)	United States	Software as a Medical Device — AI clinical tools	Require FDA clearance documentation as listing prerequisite for clinical AI; verify clearance currency	Obtain FDA clearance appropriate to AI's intended use and risk profile before commercial listing	Healthcare systems can deploy without independent FDA clearance verification; reduces procurement timeline significantly
NIST AI Risk Management Framework	United States	AI systems used in federal contexts and enterprise risk management	Require NIST AI RMF alignment documentation for enterprise tier listings	Document AI system against NIST AI RMF core functions (Govern, Map, Measure, Manage)	Standardized risk assessment framework that enterprise security teams can evaluate consistently
SOC 2 Type II	International (US standard)	Enterprise SaaS and AI services processing enterprise data	Require SOC 2 Type II for listings claiming enterprise-grade data handling	Obtain SOC 2 Type II certification from accredited auditor; maintain annually	Reduces enterprise security review requirement; SOC 2 is acceptable evidence for many enterprise procurement processes
ISO/IEC 42001 (AI Management	International	Organizations developing or using AI	Early — not yet widely required;	Implement AI Management System aligned	Emerging trust signal; expected to

Systems)			represents emerging direction for AI governance certification	with ISO 42001; pursue certification	become a significant procurement criterion as the standard matures

IP Framework for AI Marketplace Transactions

IP dimension	The question	Current legal status	Marketplace governance requirement	Commercial implication
Model output ownership	Who owns text, images, code, or other content generated by a marketplace-listed model?	Varies by model license; Copyright Office guidance: AI-only outputs not copyrightable in US; human+AI collaboration may be	Make license terms machine-readable and searchable; filter capability that commercial output rights require	Buyers who need commercial rights to AI outputs must verify license permits commercial use; marketplace should surface this proactively
Training data provenance	What data was used to train the model? Were all licenses and consent obtained?	Active litigation (Getty, NYT, GitHub Copilot); legal uncertainty about fair use for AI training	Require training data documentation as listing prerequisite; flag models with unresolved training data litigation	Enterprise buyers in sectors with significant IP risk (media, publishing, content) need provenance documentation to assess legal exposure
Fine-tuning IP	Who owns a model fine-tuned on the buyer's proprietary data using a marketplace-listed foundation model?	Varies by foundation model license; most commercial licenses assign fine-tune IP to the buyer	Standard fine-tuning terms must address: buyer IP ownership of fine-tune, provider rights (if any) in fine-tuned model, prohibition on training future base models on the fine-tune without	Buyers who fine-tune on proprietary data need clear confirmation that the resulting model is theirs to commercialize

			consent	
Derivative model rights	Can a fine-tuned model be commercially distributed? Can it be used to create another fine-tune?	Varies by license (Llama 3: commercial use permitted with revenue thresholds; many others: no commercial redistribution)	License chain tracker: when a listed model is derived from another, the original model's license constraints flow through	Determines whether a marketplace provider can commercially distribute a model derived from an open-source base; significant commercial risk if not verified before listing

Chapter Eleven — The Essentials

› EU AI Act, FDA SaMD framework, and NIST AI RMF are the three most commercially significant compliance frameworks for AI marketplace providers in 2025.

› FDA clearance requirement for clinical AI listings is the most immediately actionable compliance gate — healthcare buyers will not deploy non-cleared clinical AI regardless of marketplace trust signals.

› Training data provenance is the most significant long-term legal risk — require documentation as a listing prerequisite before litigation forces the issue.

› Fine-tuning IP terms must be explicitly specified: buyer ownership of the resulting model and prohibition on using the fine-tune in future base model training are the minimum required clauses.

› Machine-readable license terms are the infrastructure investment that reduces IP governance from a manual due diligence task to an automated filter at discovery time.

PART FIVE

Network Effects and Platform Strategy

Four network effects. Winner-takes-most dynamics. Community versus commercial.

CHAPTER TWELVE

Network Effects: Four Types That Compound

Supply-demand · Data · Integration · Trust. How they reinforce each other.

B2B AI marketplace network effects have specific dynamics that differ from consumer marketplace network effects in commercially important ways.

The B2B procurement network effect is the most B2B-specific network effect type: as more enterprise buyers in a specific industry use the marketplace to discover and procure AI capabilities, the marketplace's understanding of that industry's specific procurement requirements deepens — which attracts more industry-specific AI capability providers, which attracts more buyers, which generates more procurement data. This is distinct from the generic supply-demand flywheel because it is industry-specific and trust-anchored: an enterprise healthcare organization that discovers the marketplace has deep healthcare AI certification infrastructure and 50 HIPAA-certified AI capabilities is more likely to use the marketplace as a primary procurement channel than a generic marketplace with 5,000 uncertified capabilities.

The compliance certification network effect is the most durable B2B network effect: each certified listing makes all other listings more valuable to enterprise buyers because it demonstrates that the certification process is genuine and meaningful. When the marketplace's security certification process identifies a vulnerability in a listed product and requires remediation before the listing can be restored, it validates the certification's rigor for all other certified listings. Enterprise buyers who rely on marketplace certifications are more willing to use marketplace certifications as a procurement shortcut — reducing their per-vendor security review burden — when they have evidence that the certifications are enforced.

The procurement data network effect compounds with transaction volume: each enterprise procurement transaction generates data about the procurement requirements (what certifications were needed, what contract terms were negotiated, what integration support was required) that improves the marketplace's ability to pre-clearance future vendor listings for similar requirements. The marketplace that has processed 10,000 enterprise AI procurement transactions has accumulated the procurement pattern data that allows it to predict what a new enterprise buyer in the same industry and size cohort will require — and to pre-qualify listed capabilities against those requirements before the buyer even begins their evaluation.

The partner ecosystem network effect is specific to B2B marketplaces that develop system integrator and consulting partner channels: each certified marketplace partner creates implementation capability that reduces the enterprise buyer's risk of a failed deployment, which increases the enterprise buyer's confidence in purchasing from the marketplace, which attracts more enterprise buyers, which attracts more consulting partners seeking marketplace deployments to implement. The Salesforce AppExchange partner ecosystem — 3.7 million certified professionals — is the largest example of this network effect at work. The density of the partner ecosystem is a genuine competitive advantage that a competing marketplace cannot quickly replicate.

The winner-takes-most dynamics in B2B AI marketplaces are strong within specific vertical markets but less absolute across verticals. Within a specific industry segment (healthcare AI procurement, legal AI procurement, financial services AI procurement), the marketplace that achieves critical mass — in certified listings, in enterprise buyer relationships, and in compliance infrastructure — creates network effects that are very difficult for late entrants to overcome. But unlike consumer marketplaces where the winner takes all globally, B2B AI marketplaces can have different winners in different verticals simultaneously. The strategy implication: establish dominant positions in specific verticals rather than attempting to win all verticals simultaneously against well-capitalized horizontal competitors.

AI Marketplace Network Effects — Four Types					
Network effect type	**Mechanism**	**Strength**	**Time to establish**	**Most relevant archetype**	**Example**
Supply-demand flywheel	More providers attract more buyers (more options); more buyers attract more providers (more revenue potential)	Medium — requires critical mass before self-sustaining; vulnerable to competitive subsidy in pre-critical-mass period	12–24 months to reach sustainable flywheel at vertical scale	All archetypes	Hugging Face: more models attract developers; more developers attract model publishers; more publishers attract enterprise buyers
Data network effect	Each transaction generates data that improves discovery, evaluation, and matching for all subsequent transactions	High — compounds continuously; cannot be purchased; creates genuine intelligence advantage over competitors without comparable data	24–36 months to accumulate meaningful intelligence advantage	Model hub + cloud marketplace (transaction volume required)	Hugging Face Open LLM Leaderboard improves as more models are evaluated; recommendations improve as download patterns accumulate
Integration network effect	Each certified integration creates switching costs for buyers embedded in the integrated workflow; more integrations attract more buyers who want those specific	Very high — each integration is a specific switching cost; impossible to reverse without disrupting live operations	Varies by integration count; substantial moat at 100+ integrations	Enterprise application marketplace	Salesforce AppExchange: 7,000 integrations means each enterprise customer has multiple workflow dependencies; switching from Salesforce requires rebuilding dependencies

	integrations				
Trust network effect	Each successfully certified, monitored, and maintained listing makes all other listings more trustworthy by demonstrating that the trust infrastructure is real	Medium — positive but not self-reinforcing in the same way as the others; requires active maintenance to sustain	12–18 months to establish credibility through demonstrated enforcement	All archetypes — most critical for enterprise and healthcare marketplaces	AWS Marketplace security review: documented removals of non-compliant listings make remaining listings more trustworthy; enterprise buyers rely on the review process, not individual listing claims

Chapter Twelve — The Essentials

› Four network effects: supply-demand flywheel (medium strength), data flywheel (high, compounds continuously), integration network (very high, creates operational switching costs), trust network (medium, requires active maintenance).

› Hugging Face's compound network effect — all four types reinforcing each other — explains its near-monopoly position in ML model discovery despite zero marketing budget.

› The data network effect is the most strategically important to establish first: it creates intelligence advantages that compound and cannot be purchased once established.

› The integration network effect is the most durable once established: operational switching costs are real barriers that rational buyers do not overcome without compelling reason.

› Network effects determine winners in AI marketplace competition — the strategy question is not 'what features do we build?' but 'which network effect do we establish first, and how do we compound it?'

CHAPTER THIRTEEN

Winning the Vertical: Competitive Strategy for AI Marketplaces

Horizontal vs vertical. First-mover advantage. Community vs commercial alignment.

B2B AI marketplace network effects have specific dynamics that differ from consumer marketplace network effects in commercially important ways.

The B2B procurement network effect is the most B2B-specific network effect type: as more enterprise buyers in a specific industry use the marketplace to discover and procure AI capabilities, the marketplace's understanding of that industry's specific procurement requirements deepens — which attracts more industry-specific AI capability providers, which attracts more buyers, which generates more procurement data. This is distinct from the generic supply-demand flywheel because it is industry-specific and trust-anchored: an enterprise healthcare organization that discovers the marketplace has deep healthcare AI certification infrastructure and 50 HIPAA-certified AI capabilities is more likely to use the marketplace as a primary procurement channel than a generic marketplace with 5,000 uncertified capabilities.

The compliance certification network effect is the most durable B2B network effect: each certified listing makes all other listings more valuable to enterprise buyers because it demonstrates that the certification process is genuine and meaningful. When the marketplace's security certification process identifies a vulnerability in a listed product and requires remediation before the listing can be restored, it validates the certification's rigor for all other certified listings. Enterprise buyers who rely on marketplace certifications are more willing to use marketplace certifications as a procurement shortcut — reducing their per-vendor security review burden — when they have evidence that the certifications are enforced.

The procurement data network effect compounds with transaction volume: each enterprise procurement transaction generates data about the procurement requirements (what certifications were needed, what contract terms were negotiated, what integration support was required) that improves the marketplace's ability to pre-clearance future vendor listings for similar requirements. The marketplace that has processed 10,000 enterprise AI procurement transactions has accumulated the procurement pattern data that allows it to predict what a new enterprise buyer in the

same industry and size cohort will require — and to pre-qualify listed capabilities against those requirements before the buyer even begins their evaluation.

The partner ecosystem network effect is specific to B2B marketplaces that develop system integrator and consulting partner channels: each certified marketplace partner creates implementation capability that reduces the enterprise buyer's risk of a failed deployment, which increases the enterprise buyer's confidence in purchasing from the marketplace, which attracts more enterprise buyers, which attracts more consulting partners seeking marketplace deployments to implement. The Salesforce AppExchange partner ecosystem — 3.7 million certified professionals — is the largest example of this network effect at work. The density of the partner ecosystem is a genuine competitive advantage that a competing marketplace cannot quickly replicate.

The winner-takes-most dynamics in B2B AI marketplaces are strong within specific vertical markets but less absolute across verticals. Within a specific industry segment (healthcare AI procurement, legal AI procurement, financial services AI procurement), the marketplace that achieves critical mass — in certified listings, in enterprise buyer relationships, and in compliance infrastructure — creates network effects that are very difficult for late entrants to overcome. But unlike consumer marketplaces where the winner takes all globally, B2B AI marketplaces can have different winners in different verticals simultaneously. The strategy implication: establish dominant positions in specific verticals rather than attempting to win all verticals simultaneously against well-capitalized horizontal competitors.

AI Marketplace Competitive Positioning Matrix				
Position	**When to choose**	**Required moat**	**Critical success factor**	**Risk**
Horizontal — developer-focused (Hugging Face model)	When you can establish supply-side critical mass first; when the developer community is your primary	Data network effect (usage data improves recommendations) + integration network (library adoption)	Achieve critical mass before competitors; the winner-takes-most effect is strong at this position	You are building for a sophisticated, cost-sensitive audience who can and will bypass you if

	participant			the take rate is too high
Horizontal — enterprise-buyer-focused (AWS Marketplace model)	When you have an existing enterprise buyer relationship that creates procurement integration advantages; when your ecosystem scale is sufficient	Integration network (cloud infrastructure procurement relationship) + trust network (cloud-level security credibility)	Leverage existing buyer relationships; ensure every enterprise buyer in your ecosystem is aware of the marketplace	Cloud-locked: buyers who multi-cloud will not rely on your marketplace; you serve your ecosystem, not the full enterprise market
Vertical specialist	When a specific industry has trust requirements that generic marketplaces cannot satisfy; when you have domain relationships that seed the supply side	Trust network (domain-specific evaluation) + supply-demand flywheel within the vertical	Establish domain-specific evaluation credibility before competitors; build relationships with vertical supply leaders	Horizontal competitors will eventually build vertical trust infrastructure; window to establish the moat is 2–4 years
Infrastructure neutral	When you can provide the discovery, trust, and transaction infrastructure without being tied to a specific cloud or platform	Data network effect + trust network + transaction infrastructure efficiency	Establish before cloud providers commoditize the neutral infrastructure position; requires significant upfront investment	Highest capital requirement; most competitive position to enter; subject to disruption from cloud providers who decide to compete directly

THE COMMUNITY COMMERCIALIZATION PRINCIPLE

Never make the community subsidize the commercial tier — separate the economics, share the infrastructure

The sustainable model for community-rooted AI marketplaces: the community tier gets genuine open-source infrastructure (library access, public hosting, free evaluation tools) without being

forced to contribute to commercial revenue. The commercial tier pays for infrastructure services (compute, private storage, enterprise support) that are genuinely more expensive to provide than community tier services. The community creates value through the supply network, the data flywheel, and the trust signals that attract commercial buyers. The commercial tier captures a portion of that value through infrastructure economics, not through rent extraction from community participants. Break this model — charge the community for things they currently get free, or restrict open-source access to force commercial adoption — and the community that is the foundation of the flywheel will route around the marketplace.

Chapter Thirteen — The Essentials

› Winner-takes-most dynamics are strong in AI marketplace competition — the race to critical mass in a specific position determines commercial outcomes for a decade.

› The most accessible entry point for a new AI marketplace: a specific vertical with high trust requirements, concentrated buyer segment, and sufficient specialized supply to provide discovery value.

› The community commercialization principle: separate community economics (free infrastructure, open distribution) from commercial economics (paid infrastructure services) without extracting from the community layer.

› The horizontal developer marketplace (Hugging Face) and the horizontal enterprise marketplace (AWS, Azure) are both near-critical-mass in their specific positions — new entrants should focus on vertical specialization.

› Community trust is the foundation of marketplace trust — actions that damage community trust (license restrictions, take rate increases on community models) destroy the moat that the commercial tier depends on.

PART SIX

Agent-to-Agent Marketplace Commerce

Current state. MCP integration. Governance for machine-speed transactions.

CHAPTER FOURTEEN

A2A Commerce: Already Here, Not Yet Infrastructure

Where it is live. The transaction architecture. Building A2A-ready commercial infrastructure.

Agent-to-agent B2B commerce — automated transactions where B2B AI agents purchase AI capabilities from other AI capability providers without human involvement in individual transactions — is the commercial evolution that will most significantly change the B2B AI marketplace's transaction economics over the next five years.

The B2B nature of agent commerce is non-negotiable and must be emphasized precisely. When an AI orchestration agent purchases a specialized legal analysis capability from a B2B AI marketplace, it is not acting as an individual — it is acting as the authorized representative of a business organization. The business organization has: delegated commercial authority to the agent (through its configured spending limits), established the vendor relationship through which the agent is authorized to transact (through the organization's marketplace account), and accepted legal and financial responsibility for the agent's commercial actions (through the organization's terms of service with the marketplace). Every agent transaction is a B2B transaction. The buyer of record is always the business organization, not the AI agent.

This B2B anchoring of agent commerce has important commercial infrastructure implications. The agent identity framework must connect every agent transaction to the business organization that authorized it: the agent's JWT or credential must include the organizational account identifier, the agent's spending authority within that organization, and the capability categories the agent is authorized to purchase. The marketplace's billing infrastructure must aggregate all agent transactions under the organizational account, generating the per-period invoice that the enterprise procurement system expects. The audit trail must be organized by organizational account and be queryable by organizational finance and compliance teams — not just by the agent that generated the transactions.

The four B2B governance requirements for agent commerce:

Commercial authority governance: each B2B AI agent must have a defined spending authority — the maximum amount it can commit on behalf of the organization per transaction, per day, and per period. The authority is delegated by the organization's FinOps or procurement function, recorded in the marketplace's agent registry, and enforced by the marketplace's commercial middleware at each transaction. Transactions that exceed the agent's authority must trigger a human escalation workflow rather than proceeding automatically.

Vendor relationship governance: B2B agent commerce must occur within pre-approved vendor relationships. An enterprise organization that has not completed its vendor qualification process for a specific AI capability provider cannot allow its AI agents to transact with that provider automatically. The marketplace must enforce this requirement by allowing organizations to configure approved vendor lists for each agent — agents can only transact with marketplace providers that have been pre-approved in the organizational configuration.

Data handling governance: when a B2B AI agent submits data to an AI capability provider in the course of a transaction, the data handling must comply with the organization's DPA with that provider. The marketplace must enforce data handling rules at the transaction layer — an agent cannot submit data classified as HIPAA-protected to a provider that is not under a BAA with the organization, even if the agent is technically authorized to purchase that provider's capabilities for non-protected data use cases.

Attribution and accountability governance: every agent-executed B2B marketplace transaction must be attributable to a specific business purpose — a specific project, workflow, or organizational initiative that the transaction supports. Attribution is required for internal cost accounting (which budget does this charge against?), for compliance reporting (what AI capabilities did we use, for what purposes, in this compliance period?), and for audit response (when auditors ask what AI tools the

organization used for a specific process, the transaction log must provide a complete answer).

A2A Marketplace Commerce Maturity by Vertical					
Vertical	**Current A2A transaction pattern**	**Commercial maturity**	**Missing infrastructure**	**Timeline to full A2A marketplace**	**Key players building it**
Research and literature	AI agents querying Semantic Scholar, PubMed, arXiv, Crossref APIs autonomously within research workflows	Pre-commercial — de facto A2A transactions through free/subscription APIs; no per-transaction commercial settlement	Real-time capability discovery (capabilities are hardcoded, not discovered at runtime); per-use commercial settlement	2–3 years to first commercial A2A marketplace transactions in research	Elsevier, Springer, academic library consortia exploring commercial API models
Software development	AI coding agents invoking code analysis, security scanning, test generation, documentation tools through orchestration frameworks	Early commercial — LangChain tool libraries are de facto marketplaces; commercial settlement not yet standard	Dynamic capability selection (agents choose from alternatives at runtime); micropayment settlement	1–2 years — closest to production-ready A2A marketplace commerce	Snyk, Semgrep, Codacy building tool APIs; LangChain/LlamaIndex building the orchestration layer
Data and analytics	AI analytics agents purchasing specialized data enrichment, market data, and	Partial commercial — some data providers offer API access with per-query pricing; full	Capability manifest standard (so agents can discover what data products	2–3 years for standardized A2A data marketplace	Bloomberg, Refinitiv, Dun & Bradstreet, FactSet exploring agent-accessible data product APIs

	domain-specific calculations	A2A discovery not yet standard	exist); automatic contract execution		
Legal and professional services	AI agents invoking contract analysis, regulatory research, case law search, and compliance checking tools	Early commercial — Harvey AI, Westlaw, Lexis APIs are accessible but not A2A-discovery-compatible	Performance attestation (current accuracy metrics); automatic vendor selection based on requirements	3–5 years — regulated nature of legal AI slows commercial infrastructure standardization	Thomson Reuters, LexisNexis, Harvey AI building toward A2A-compatible professional AI service APIs
Healthcare and clinical	AI agents invoking clinical NLP, imaging analysis, drug interaction checking, clinical guideline retrieval	Pre-commercial — significant regulatory barriers to autonomous agent commercial transactions in clinical contexts	All A2A commerce infrastructure plus regulatory framework for autonomous agent clinical AI procurement	5–7 years — FDA and healthcare regulatory frameworks must evolve before autonomous agent clinical AI purchasing is viable	Epic, Cerner, healthcare-specific AI companies building toward FDA-compliant agent-accessible APIs

Chapter Fourteen — The Essentials

› A2A marketplace commerce is live in specific verticals (software development, research) and moving toward commercial maturity in others (data, legal).

› The transition from de facto A2A (hardcoded API calls) to true A2A marketplace (dynamic discovery + commercial settlement) requires three infrastructure layers: agent identity, real-time capability discovery, and micropayment settlement.

› Software development is the first vertical to reach A2A marketplace commerce maturity — LangChain/LlamaIndex tool libraries are the discovery infrastructure; commercial settlement is the remaining missing piece.

› Healthcare and legal are 5–7 years from full A2A marketplace commerce — regulatory frameworks must evolve before autonomous agent purchasing of clinical and legal AI is viable.

› The marketplace that builds A2A commerce infrastructure in software development first creates the template that other verticals will adopt — vertical-by-vertical expansion is the growth path.

CHAPTER FIFTEEN

MCP Integration and the Marketplace Protocol Stack

How MCP servers become marketplace listings. The commercial wrapper. The Hugging Face MCP integration.

The Model Context Protocol (MCP), developed by Anthropic and introduced in Book 4, is the technical standard that governs how AI agents interact with external tools and capabilities. The integration between MCP servers and B2B AI marketplaces creates the technical infrastructure for B2B agent commerce — with the B2B governance requirements described in the previous chapter implemented at the protocol layer.

The B2B MCP marketplace integration model works as follows. A B2B AI capability provider creates an MCP server that exposes their capabilities as structured tools — with defined input schemas, output schemas, and capability manifests. The B2B AI marketplace wraps this MCP server with a commercial middleware layer that enforces B2B governance requirements: organizational account validation (the calling agent is registered to a valid organizational account), spending authority verification (the transaction does not exceed the agent's configured authority), vendor relationship verification (the organizational account has a valid commercial relationship with this provider), data handling enforcement (the submitted data complies with the DPA between the organization and this provider), and transaction logging (the transaction is recorded in the organizational account's audit log with full attribution).

The MCP tool registry — the B2B marketplace's catalog of available MCP-exposed AI capabilities — is the discovery layer for agent-driven B2B commerce. An AI

orchestration agent that needs to assemble a multi-step workflow queries the MCP tool registry to identify available capabilities by capability type, interface compatibility, performance characteristics, and commercial availability. The registry returns a structured list of candidate capabilities with their manifests, pricing, and performance data — allowing the agent to select the optimal capability for its current task programmatically.

The commercial layer that B2B marketplace operators must add to MCP integration goes beyond what the open-source MCP protocol specifies. MCP defines the technical interaction between agents and tools — the message formats, the error handling, the streaming patterns. It does not define the commercial layer — the pricing, the settlement, the authority governance, or the compliance enforcement. B2B marketplace operators must build this commercial layer as a wrapper around the technical MCP protocol, making B2B commercial requirements transparent to the AI agent's task logic while enforcing them at the infrastructure layer.

The Hugging Face MCP integration launched in 2025 — which makes every Hugging Face Space available as an MCP server — represents the first significant deployment of marketplace-connected MCP at ecosystem scale. In its current form, the commercial layer is missing: MCP invocations of Hugging Face Spaces are free or covered by existing subscriptions, with no per-invocation B2B commercial settlement. The next phase — adding B2B commercial settlement, organizational account enforcement, and spending authority governance to the MCP integration layer — is the infrastructure investment that converts Hugging Face MCP from a developer tool into a B2B agent commerce platform.

MCP Marketplace Integration — Commercial Layer Design			
MCP component	**Commercial layer addition**	**What it enables**	**Implementation complexity**
Tool definition (name, description, input schema,	Add pricing annotation: price per call, pricing model type, billing unit	Agent can evaluate cost of tool invocation before calling; enables	Low — JSON annotation to existing tool manifest

output schema)		price-aware capability selection in cost-constrained workflows	
Tool invocation	Intercept with commercial middleware: validate agent authentication, check balance/authority, log invocation, deduct cost	All A2A commercial requirements satisfied at the invocation layer transparently to the tool implementation	Medium — requires marketplace-operated MCP proxy that wraps the provider's MCP server
Tool response	Attach transaction receipt to response: transaction_id, amount_charged, remaining_balance	Creates the audit trail required for enterprise billing reconciliation and dispute resolution	Low — append to response metadata
Error handling	Distinguish commercial errors (insufficient balance, authority limit exceeded) from technical errors (tool failure)	Enables appropriate agent behavior: retry on technical error, request human authorization on commercial error	Low — add commercial error codes to standard error handling
Provider metrics	Add real-time performance metrics to tool manifest: latency_p50, latency_p95, error_rate, last_updated	Enables agent to make quality-aware tool selection; choose fastest/most reliable option at runtime	Medium — requires provider to expose metrics API that marketplace aggregates and publishes

Chapter Fifteen — The Essentials

› MCP servers become marketplace listings by adding a commercial wrapper that intercepts invocations to enforce pricing, authentication, logging, and settlement.

› The commercial wrapper is transparent to the tool implementation — providers do not need to change their MCP server; the marketplace proxy handles the commercial layer.

› Hugging Face's 2025 MCP integration represents the first significant step toward A2A marketplace commerce at ecosystem scale — every Space becomes a potential AI agent tool with marketplace discovery.

› The remaining gap: commercial settlement. MCP invocations are currently free or covered by subscription — per-invocation commercial settlement is the next infrastructure requirement.

› The protocol stack: MCP defines how agents use capabilities; the Dynamic AI Capability Pricing Protocol (Chapter 9) defines how they buy them — both are required for full A2A marketplace commerce.

CHAPTER SIXTEEN

Governing Agent Commerce: Liability, Spending Controls, and Anti-Manipulation

Who is responsible when agents transact. Enterprise spending governance. Marketplace manipulation prevention.

Agent-to-agent B2B commerce — automated transactions where B2B AI agents purchase AI capabilities from other AI capability providers without human involvement in individual transactions — is the commercial evolution that will most significantly change the B2B AI marketplace's transaction economics over the next five years.

The B2B nature of agent commerce is non-negotiable and must be emphasized precisely. When an AI orchestration agent purchases a specialized legal analysis capability from a B2B AI marketplace, it is not acting as an individual — it is acting as the authorized representative of a business organization. The business organization has: delegated commercial authority to the agent (through its configured spending limits), established the vendor relationship through which the agent is authorized to transact (through the organization's marketplace account), and accepted legal and financial responsibility for the agent's commercial actions (through the organization's terms of service with the marketplace). Every agent transaction is a B2B transaction. The buyer of record is always the business organization, not the AI agent.

This B2B anchoring of agent commerce has important commercial infrastructure implications. The agent identity framework must connect every agent transaction to the business organization that authorized it: the agent's JWT or credential must include the organizational account identifier, the agent's spending authority within that organization, and the capability categories the agent is authorized to purchase. The marketplace's billing infrastructure must aggregate all agent transactions under the organizational account, generating the per-period invoice that the enterprise procurement system expects. The audit trail must be organized by organizational account and be queryable by organizational finance and compliance teams — not just by the agent that generated the transactions.

The four B2B governance requirements for agent commerce:

Commercial authority governance: each B2B AI agent must have a defined spending authority — the maximum amount it can commit on behalf of the organization per transaction, per day, and per period. The authority is delegated by the organization's FinOps or procurement function, recorded in the marketplace's agent registry, and enforced by the marketplace's commercial middleware at each transaction. Transactions that exceed the agent's authority must trigger a human escalation workflow rather than proceeding automatically.

Vendor relationship governance: B2B agent commerce must occur within pre-approved vendor relationships. An enterprise organization that has not completed its vendor qualification process for a specific AI capability provider cannot allow its AI agents to transact with that provider automatically. The marketplace must enforce this requirement by allowing organizations to configure approved vendor lists for each agent — agents can only transact with marketplace providers that have been pre-approved in the organizational configuration.

Data handling governance: when a B2B AI agent submits data to an AI capability provider in the course of a transaction, the data handling must comply with the organization's DPA with that provider. The marketplace must enforce data handling rules at the transaction layer — an agent cannot submit data classified as HIPAA-

protected to a provider that is not under a BAA with the organization, even if the agent is technically authorized to purchase that provider's capabilities for non-protected data use cases.

Attribution and accountability governance: every agent-executed B2B marketplace transaction must be attributable to a specific business purpose — a specific project, workflow, or organizational initiative that the transaction supports. Attribution is required for internal cost accounting (which budget does this charge against?), for compliance reporting (what AI capabilities did we use, for what purposes, in this compliance period?), and for audit response (when auditors ask what AI tools the organization used for a specific process, the transaction log must provide a complete answer).

Enterprise A2A Spending Governance Architecture

Governance layer	What it controls	Implementation	Commercial implication
Organization budget	Total monthly/quarterly AI marketplace spend across all agents	Token Budget Level 1 (Book 3a): organization-wide limit; marketplace spending contributes to the same budget as model API spending	CFO visibility into total AI marketplace spend; quarterly variance analysis against approved budget
Function/team budget	AI marketplace spend by business unit or team (legal AI, engineering AI, finance AI)	Token Budget Level 2: function-level allocation within organization budget; FinOps team manages cross-function trade-offs	Department head accountability for AI marketplace spend; enables cost attribution for ROI analysis by function

Project/workflow budget	AI marketplace spend for a specific AI project or recurring workflow	Token Budget Level 3: workflow-specific allocation; enables project-level cost tracking	Project manager accountability; enables milestone-based budget review for multi-phase AI projects
Agent budget	Maximum spend authority for a specific AI agent in a single interaction or day/week period	Token Budget Level 4: per-agent limit enforced at the marketplace commercial middleware layer	Prevents individual agents from exceeding their configured commercial authority; automatic human escalation at limit
Capability category restriction	Which types of capabilities each agent is authorized to purchase (research AI, code AI, but not legal AI without human approval)	Capability category allow/deny list per agent identity in marketplace configuration	Governs what AI agents can purchase autonomously vs what requires human approval at the category level

Chapter Sixteen — The Essentials

› A2A marketplace governance requires four layers of spending control: organization, function, project, and agent — matching the Token Budget hierarchy from Book 3a.

› Liability in automated transactions: the deploying enterprise is ultimately responsible for agent actions; marketplace provides the audit trail required for dispute resolution.

› Anti-manipulation: agents optimizing for cost efficiency or task completion rate will find unexpected optima — marketplace governance must anticipate these optimization patterns in the commercial infrastructure design.

› Capability category restrictions — allowing some categories of autonomous purchasing while

requiring human approval for others — is the practical governance mechanism for enterprise AI deployment.

› The audit log is the enterprise governance foundation: every agent invocation must be logged with agent identity, capability invoked, parameters, and cost — the basis for all spending accountability.

PART SEVEN

Building and Operating the Marketplace

Build vs partner. Revenue architecture. The operational model.

CHAPTER SEVENTEEN

Build, Partner, or List: The Strategic Decision

When to build a dedicated marketplace. When to partner. The hybrid approach.

Building a B2B AI marketplace requires simultaneous investment in supply, demand, and trust infrastructure — with a specific sequencing that differs from consumer marketplace development because B2B trust requirements are gates, not features.

In consumer marketplaces, trust infrastructure can be added progressively as the platform scales — start with the supply and demand flywheels, add trust mechanisms as problems arise. In B2B marketplaces, especially enterprise-focused B2B marketplaces, trust infrastructure is the prerequisite for enterprise buyer adoption, not a progressive addition. Enterprise IT teams will not discover and evaluate AI capabilities through a marketplace that does not demonstrate baseline security and compliance infrastructure before the evaluation begins.

The B2B AI marketplace build sequence:

Phase 1 — Trust infrastructure foundation (months 1–9): Build the certification program before acquiring listings. Define the security requirements, compliance documentation requirements, and performance attestation standards that all listed capabilities must meet. Build the vendor certification process and the certification verification infrastructure. Recruit the first 20–30 certified capability providers who meet the standard — even if the certification process is partially manual in Phase 1. The goal of Phase 1 is to have a small, trusted catalog ready before opening to enterprise buyer traffic.

Phase 2 — Enterprise buyer discovery and procurement infrastructure (months 6–18): Build the discovery infrastructure (taxonomy, search, evaluation tools) for enterprise buyers, using the certified catalog from Phase 1 as the inventory. Build the procurement support infrastructure (procurement document packages, private offer infrastructure, enterprise account management tools). Identify and engage the first 10–20 enterprise buyer organizations who will serve as design partners — their feedback on the discovery and procurement experience is the design input for Phase 2 infrastructure.

Phase 3 — Retail B2B channel activation (months 12–24): Once enterprise-grade trust infrastructure is established and tested with design partner buyers, open the self-serve retail B2B channel. The retail B2B buyer discovers and evaluates on the same platform as the enterprise buyer, but through a simplified self-serve flow with standard commercial terms and automated account provisioning. The enterprise trust infrastructure that was built for the enterprise buyer is also the trust signal that makes the retail B2B buyer confident in purchasing without extended evaluation.

Phase 4 — Agent commerce infrastructure (months 18–36): Build the machine-readable capability manifests, real-time pricing APIs, performance attestation endpoints, and B2B organizational account settlement infrastructure required for agent-driven B2B commerce. Partner with AI orchestration framework developers (LangChain, LlamaIndex, Claude Code, AutoGen) to integrate the marketplace's capability registry into their tool discovery systems. The agent commerce infrastructure serves the agent-

buyer customer segment and creates the automated transaction volume that grows beyond human-initiated marketplace transactions at scale.

Phase 5 — Partner channel activation (months 24+): Build the partner portal and channel program for system integrators and consulting organizations. The partner channel is particularly important for enterprise B2B AI deployment because many enterprise organizations lack the internal AI implementation capability to deploy marketplace-sourced AI independently — they need a certified implementation partner. The marketplace's partner certification program and deal registration system creates the commercial infrastructure for the partner channel.

Build vs Partner vs List — Decision Framework

Decision	When to choose	Capital requirement	Time to market	Competitive risk
Build a dedicated marketplace	You have a specific vertical with high trust requirements not served by existing marketplaces; you have supply-side relationships to seed the catalog; your competitive advantage is in trust/evaluation infrastructure for the specific vertical	High — $5M–$20M to reach viable MVP for enterprise vertical; $50M+ to reach critical mass	18–36 months to viable enterprise marketplace in a well-defined vertical	High execution risk; moderate competitive risk if you establish the vertical trust moat before competitors
Build on top of existing marketplace (become a specialized layer)	You can provide a vertical-specific curation, evaluation, and trust layer on top of a general marketplace; you add value through domain expertise, not discovery infrastructure	Medium — $1M–$5M to build the trust and curation layer; less infrastructure investment required	6–12 months to vertical-specific trust layer on existing discovery infrastructure	Dependency on underlying marketplace; platform risk if marketplace changes terms; lower moat than building the full stack

Partner with existing marketplace (CPPO / co-sell)	You are an AI capability provider seeking enterprise distribution; the marketplace's buyer base overlaps your target customers; the take rate is justified by the distribution value	Low — primarily commercial negotiation and integration cost	2–6 months for listing + certification; immediate access to marketplace buyer base	Marketplace dependency; take rate competes with your margin; limited differentiation from other listed providers
List on multiple marketplaces (multi-channel)	You are an AI capability provider; different buyer segments use different marketplaces; you want distribution diversification	Low — incremental integration per marketplace; primarily listing and certification cost	Varies by marketplace; sequential, not parallel listing recommended for quality management	Reduced dependency on any single marketplace; complexity of managing multiple listing standards; potential take rate differences across marketplaces

Chapter Seventeen — The Essentials

› The build/partner/list decision depends on whether your competitive advantage is in marketplace infrastructure (build) or in AI capabilities (partner/list).

› Building a dedicated vertical marketplace is the highest-capital, highest-risk, highest-reward option — justified when a specific vertical's trust requirements create a genuine evaluation moat.

› The CPPO model on AWS Marketplace is the most commercially effective listing approach for enterprise AI providers — it combines marketplace distribution with custom commercial terms.

› Multi-channel listing (multiple marketplaces) is the distribution diversification strategy for established AI providers — reduces single-marketplace dependency while maintaining broad reach.

› The hybrid approach — list on existing marketplaces while building direct developer community in parallel — is the scale strategy that reduces long-term marketplace dependency as the direct channel matures.

CHAPTER EIGHTEEN

Revenue Architecture and Operational Model

Five revenue streams. The platform P&L. Operating a two-sided marketplace at scale.

The revenue architecture for a B2B AI marketplace operator reflects the B2B commercial dynamics described throughout this book — with the specific feature that B2B revenue concentrates in fewer, larger transactions than consumer marketplace revenue, requiring commercial infrastructure that supports both high-volume retail B2B transactions and low-volume, high-value enterprise B2B transactions.

Revenue stream 1 — Retail B2B subscription take rate: the marketplace charges a percentage (12–15%) of standard subscription fees paid by retail B2B buyers for listed capabilities. This is the highest-volume, lowest-average-value revenue stream. For a marketplace with 10,000 retail B2B subscribers paying an average of $500/month for listed capabilities, and a 13% take rate, this stream generates $650K/month in marketplace revenue. The retail B2B take rate is the commercial engine that funds platform operations while enterprise revenue concentrates.

Revenue stream 2 — Enterprise B2B transaction commission: the marketplace charges a lower percentage (8–10%) of negotiated enterprise contract values executed through the marketplace's Private Offer or equivalent infrastructure. A marketplace with 100 active enterprise B2B contracts averaging $200K/year, at a 9% take rate, generates $1.8M/year from this stream — from 100 transactions rather than 10,000. The enterprise commission is the highest-absolute-value revenue stream despite the lower take rate.

Revenue stream 3 — Certification and compliance services: the marketplace charges vendors for the certification process that is prerequisite for listing. Security certification: $5,000–$25,000 per certification depending on depth. Domain performance attestation: $10,000–$50,000 per domain benchmark evaluation depending on complexity. Annual certification renewal: 40–60% of initial certification cost. A

marketplace with 500 certified listings generating average $15K initial certification and $8K annual renewal generates $7.5M from initial certifications and $4M/year from renewals.

Revenue stream 4 — Procurement infrastructure subscriptions: the marketplace charges enterprise buyer organizations for premium procurement features — enhanced compliance documentation, dedicated procurement support, custom SLA monitoring, and priority dispute resolution. Enterprise procurement subscriptions at $50,000–$200,000/year per enterprise organization serve the largest and most procurement-intensive enterprise buyers.

Revenue stream 5 — Agent commerce transaction processing: for B2B agent-to-agent transactions executed through the marketplace's commercial middleware, the marketplace charges a transaction processing fee (smaller than the standard take rate — typically 2–5% — reflecting the automated nature of the transaction and the lower support cost). As agent commerce volume grows, this stream becomes increasingly significant — 1,000 agent-initiated B2B transactions per day at an average value of $5 and a 3% processing fee generates $150/day initially, scaling to significant revenue as agent commerce volume multiplies.

The B2B marketplace P&L has a specific structure: high fixed costs (certification program, compliance infrastructure, enterprise account management) that justify the trust premium, and revenue that scales with both transaction volume (retail B2B) and transaction value (enterprise B2B). The marketplace achieves sustainable unit economics when the certification infrastructure is amortized across enough certified listings, and the enterprise revenue base is large enough to fund the compliance and procurement support infrastructure at appropriate quality levels.

AI Marketplace Revenue Architecture — Five Streams				
Revenue stream	**Pricing model**	**Gross margin**	**Scale with network**	**Primary beneficiary**
Infrastructure services	Per-compute-hour, per-GB	60–80% — infrastructure	Linear with transaction	Platform — anchored to real infrastructure

	storage, per-API-call for hosting and serving capabilities	cost is the primary variable; margin expands with scale economies	volume; scales independently of network effect	costs that providers would otherwise bear; justifiable against community critique
Transaction fees (take rate)	Percentage of commercial transaction value (5–25% depending on tier and value-add)	High — primarily revenue less payment processing; minimal variable cost per transaction	Scales super-linearly with GMV growth (volume × AOV)	Platform — requires genuine value-add to be sustainable; subject to bypass if unjustified by value
Certification and compliance services	Per-listing fee for security review, compliance certification, domain attestation ($5K–$50K per certification)	Very high — primarily professional time and testing infrastructure; significant scale economies	Scales with listing growth; less correlated with transaction volume	Platform + Providers — providers receive a trust signal that unlocks enterprise buyers; platform receives certification revenue
Data and analytics products	Subscription or per-report for market intelligence, usage insights, benchmark comparison data	High — data products have near-zero marginal cost once created	Super-linear with data accumulation — more transactions create richer intelligence that commands higher prices	Platform primarily; providers benefit from market intelligence they could not generate alone
Developer tooling subscriptions	Per-user or per-organization subscription for SDKs, testing environments, integration templates	Medium — requires ongoing tooling investment; improves with scale economies	Scales with active developer community; grows faster than transaction volume as developer ecosystem deepens	Providers primarily — tooling reduces the cost of building high-quality marketplace listings; platform earns from the provider's investment in listing quality

Chapter Eighteen — The Essentials

› Five revenue streams with different margin profiles: infrastructure services (60–80%), transaction fees (very high), certification services (very high), data products (high), developer tooling (medium).

› Infrastructure services are the most commercially defensible stream — anchored to real costs, not subject to bypass criticism, and scale with platform growth.

› Certification and compliance services are the highest-margin stream when combined with genuine trust infrastructure — they are paid for the value they unlock (enterprise procurement access), not for the effort required to produce them.

› Data and analytics products are the compounding revenue stream — they become more valuable (and command higher prices) as the platform accumulates more transaction data.

› The platform P&L should track all five streams separately — each has different growth dynamics, different margin profiles, and different strategic importance.

EXTENDED CASE STUDIES

Four Marketplace Archetypes in Depth

Hugging Face · Replicate · AWS Marketplace · Salesforce AppExchange

CASE STUDY A

Hugging Face: The Community Flywheel at Ecosystem Scale

From chatbot startup to the most important infrastructure company in the AI economy. The commercial architecture that made it sustainable.

Hugging Face's B2B commercial evolution — from community hub to enterprise AI marketplace — illustrates how a developer-focused platform navigates the transition to

genuine B2B commercial infrastructure without destroying the community that created the platform's value.

The B2B commercial trajectory: Hugging Face launched as a developer community platform. Its initial commercial transaction type was "free" — the community model was based on open access with no commercial relationship. The first B2B commercial product was the Pro subscription at $9/month — a retail B2B product accessible through self-serve signup. The Enterprise Hub — with private repositories, SSO integration, enterprise support, and organizational management — was the first enterprise B2B product. The Inference Endpoints product — dedicated compute for production AI deployments with enterprise SLAs — is the enterprise B2B product that most directly addresses enterprise IT security and reliability requirements.

The B2B customer segment differentiation in Hugging Face's commercial architecture reflects precisely the retail vs enterprise B2B distinction described in this book. Retail B2B customers use the Pro or Team tier through self-serve signup: these are engineering teams at businesses that need private model repositories, higher compute allocation, and organizational collaboration features without requiring IT-led procurement. Enterprise B2B customers use Enterprise Hub through a negotiated commercial relationship: these are large organizations with IT security requirements (SSO, audit logs, compliance reporting), organizational governance requirements (team management, access controls, spend tracking), and commercial relationship requirements (contractual SLAs, dedicated support, custom DPA terms).

The same underlying platform serves both segments. The retail B2B customer sees a streamlined self-serve interface with standard commercial terms. The enterprise B2B customer sees an enterprise-configured environment with their organization's SSO, their approved model catalog, their spend governance controls, and their compliance documentation. Both are B2B customers. The channel and the commercial infrastructure differ; the platform is the same.

The 2025 MCP integration's B2B commercial implications: Hugging Face Spaces exposed as MCP servers creates the discovery layer for B2B agent commerce. The

missing commercial layer — B2B organizational account settlement, spending authority governance, vendor relationship enforcement — is the infrastructure investment that converts the technical MCP capability into a B2B agent commerce platform. When Hugging Face adds B2B commercial governance to the MCP integration layer, every Space becomes not just a tool that Claude can use, but a B2B purchasable capability that enterprise AI agents can discover, evaluate, and transact with under their organization's commercial governance.

CASE STUDY: HUGGING FACE *Model Hub Archetype — Community-First Commercial Architecture*	
The flywheel	Transformers library adoption → community publishes models → downloads accumulate quality signals → Open LLM Leaderboard surfaces best models → enterprise buyers discover → enterprise buys infrastructure → revenue funds platform development → Transformers library improves → restart.
The commercial architecture	Free tier: unlimited public model hosting, free downloads, community evaluation tools — zero take rate on community models. Commercial tier: Inference API ($0.06/1M tokens), Inference Endpoints (from $0.06/hour), Enterprise Hub (custom pricing). Revenue is infrastructure economics, not take-rate rent extraction.
The open-source moat	Transformers library: open-source Python library providing access to 300,000+ models through a single interface. Datasets library: open-source library for 50,000+ datasets. Evaluate library: open-source evaluation framework. These three libraries make Hugging Face the infrastructure dependency for the entire ML practitioner community — removing the dependency would require replacing three ubiquitous tools simultaneously.
The A2A commerce frontier	2025 MCP integration: every Hugging Face Space exposed as an MCP server, discoverable and invocable by Claude and other MCP-compatible AI systems. Commercial settlement for MCP invocations: in development. The platform is building the discovery infrastructure before the commercial infrastructure — the same sequencing that made the original model hub successful.
The numbers	$70M+ ARR (estimated 2024). $4.5B valuation (2023 funding round). 1M+ public models. 500,000+ datasets. 350,000+ Spaces. Backed by Google, Nvidia, Amazon, Salesforce, and others — the strategic investors reflect the platform's centrality to the AI infrastructure stack.

CASE STUDY B

Replicate: Per-Inference Infrastructure for the Open-Source Ecosystem

Developer-focused inference infrastructure. Fine-tuning as a marketplace. The A2A commerce readiness.

AWS Marketplace's B2B commercial model is the most instructive case study in how procurement integration creates enterprise B2B marketplace dominance — and why the B2B procurement dynamics of a cloud provider marketplace differ fundamentally from the consumer or developer marketplace model.

The B2B procurement integration as the primary moat: AWS Marketplace's competitive advantage is not discovery superiority or trust infrastructure leadership — it is the procurement integration that allows enterprise B2B buyers to add AI service subscriptions to their existing AWS commercial relationship without new vendor onboarding, new security reviews, new procurement approvals, or new invoice relationships. An enterprise technology organization that is already spending $50M annually with AWS has a single vendor relationship covering all cloud infrastructure. Adding an AI service from AWS Marketplace requires no additional relationship — it is additive spend within an existing vendor relationship.

The B2B procurement cycle reduction this creates is commercially significant. An enterprise AI team that would otherwise spend 3–4 months in vendor onboarding (legal review, security assessment, contract negotiation, procurement approval) for a new AI service vendor can complete the procurement in days through AWS Marketplace. The 3–4 month cycle reduction is worth meaningful commercial value to enterprise AI teams operating in competitive markets where time-to-deployment is a strategic advantage.

The Private Offer model for enterprise B2B AI: the CPPO (Channel Partner Private Offer) structure allows AI service vendors to create custom commercial terms for specific enterprise AWS customers — terms that include volume discounts, custom SLAs, enterprise-specific compliance commitments, and negotiated liability terms — while

executing the transaction through the enterprise customer's existing AWS commercial relationship. The enterprise buyer approves the custom offer in their AWS console, the subscription appears on their monthly AWS bill, and the AI service vendor receives payment through AWS Marketplace settlement (minus the marketplace take rate). The enterprise B2B procurement convenience is preserved while the commercial terms are customized for the enterprise relationship.

The B2B agent commerce evolution at AWS Marketplace: as enterprise AI agents begin purchasing AI services autonomously, the AWS Marketplace is the most commercially mature infrastructure for B2B agent commerce in the enterprise segment. Enterprise AI agents operating within AWS infrastructure can be configured to transact through AWS Marketplace using the organization's existing commercial relationship, with spending governance enforced through AWS account-level budget controls and AWS Organizations policies. The organizational account governance that enterprises already manage for their cloud infrastructure spend can be extended to govern AI agent marketplace purchasing.

CASE STUDY: REPLICATE *Inference Infrastructure Archetype — Developer-First Commercial Model*	
The commercial model	Per-second pricing for every model run on Replicate's infrastructure. Transparent, consistent pricing published per model. Fine-tuning marketplace: providers create and publish fine-tuned models; Replicate charges for training compute and running inference on the resulting models.
The A2A readiness	Replicate's per-inference pricing, REST API, and documented performance metrics make it structurally compatible with A2A marketplace commerce. A code-generating agent that needs a specific image processing capability can query Replicate's pricing API, compare performance metrics, and execute the transaction programmatically — the commercial infrastructure is already closer to A2A-ready than most marketplace platforms.
The fine-tuning economy	The most commercially interesting Replicate dynamic: a creator fine-tunes Stable Diffusion on their artistic style, publishes the fine-tune, and other developers pay per-inference to use that style. The creator earns revenue from the marketplace; the downstream developer accesses a specialized capability without building fine-tuning infrastructure. This is the composition marketplace in action — the marketplace value is in the specialization layer on top of open-

	source bases.
The competitive position	Replicate competes with Hugging Face Inference API, fal.ai, Modal, and others for inference infrastructure. The competitive dynamic: all are building on open-source models with comparable capabilities. Differentiation comes from developer experience, pricing, latency, and fine-tuning marketplace breadth — not from proprietary models. The winner is likely the platform with the best developer tooling and the most active fine-tuning community.

CASE STUDY C

AWS Marketplace: Enterprise Distribution Through Procurement Integration

How cloud provider marketplaces win enterprise AI distribution by solving procurement friction, not discovery.

ServiceNow Store illustrates a B2B AI marketplace model that is distinct from both the developer hub and the cloud provider marketplace archetypes — the platform-embedded enterprise application marketplace, where the buyer's procurement decision is inseparable from their existing platform commitment.

The B2B procurement context: ServiceNow customers are enterprise organizations that have made a multi-year, multi-million-dollar commitment to the ServiceNow platform for enterprise workflow automation. When a ServiceNow customer evaluates an AI capability in the ServiceNow Store, they are not making an independent AI procurement decision — they are making a decision about extending their existing ServiceNow investment with AI-powered capabilities. The B2B procurement dynamics are fundamentally different from standalone AI procurement: the vendor relationship (ServiceNow) already exists, the security review is already done (ServiceNow's certification covers the platform; the Store application security review extends it), and the procurement approval process is simplified (it is an extension of the existing ServiceNow contract, not a new vendor relationship).

The ServiceNow AI Trust Layer: ServiceNow's AI governance framework for Store-listed AI applications requires that AI-generated outputs be clearly labeled, that customer data not be used to train foundation models without explicit consent, and that AI outputs be traceable to their sources. For B2B enterprise buyers in regulated industries (healthcare, financial services, government), the AI Trust Layer provides the governance documentation that enterprise compliance teams require — without each buyer having to independently assess each AI application's governance posture.

The B2B commercial model for ServiceNow Store AI applications: ISVs building AI applications for ServiceNow face a specific B2B commercial challenge — their application must justify additional licensing fees to customers who are already paying significant ServiceNow platform fees. The commercial structures that work in this context are: outcome-based components (the AI application includes SLA commitments for specific workflow automation outcomes, with financial remedies for performance failures), workflow-specific pricing (pricing tied to the number of automated workflows rather than to seat count or API calls), and integration-based expansion (the application includes connectors to multiple external data sources, with pricing reflecting the expanded scope of automation).

CASE STUDY: AWS MARKETPLACE *Cloud Provider Enterprise Marketplace — Procurement Integration as the Moat*	
The procurement moat	Enterprise buyers with $50M+ annual AWS spend can add AI services to their existing AWS invoice — no new vendor relationship, no new procurement approval, no new security review (beyond marketplace listing review). This procurement integration is worth 3–6 months of enterprise sales cycle friction — the most significant distribution advantage in the enterprise AI marketplace.
The Private Offer model	CPPO (Channel Partner Private Offer) allows AI vendors to create custom commercial terms for specific enterprise customers within the Marketplace framework. Enterprise sees the custom offer in their AWS account, accepts the custom terms, pays through existing AWS billing. The AI vendor gets procurement-friction-free enterprise distribution; the enterprise gets custom terms within their existing vendor relationship.
The competitive limitation	AWS Marketplace effectiveness is limited to AWS-native buyers. Multi-cloud enterprises, Azure-native organizations, and companies that have not

	standardized on AWS infrastructure get limited value from AWS Marketplace procurement integration. The addressable market is large but not universal — approximately 33% of enterprise cloud spending is on AWS, which is the effective market ceiling for AWS Marketplace's procurement integration moat.
The AI category growth	AWS Marketplace's AI and Machine Learning category is one of its fastest-growing categories. The combination of enterprise cloud infrastructure purchasing and the procurement friction elimination creates a natural channel for AI service companies targeting enterprise buyers who are already in the AWS ecosystem.

CASE STUDY D

Salesforce AppExchange: The Mature Enterprise Marketplace Archetype

18 years of marketplace development. $5B+ in cumulative transactions. The template for enterprise AI application marketplaces.

A purpose-built vertical B2B AI marketplace for the legal sector illustrates what the enterprise vertical specialist archetype looks like in practice — and why the trust infrastructure investment required to serve a regulated B2B vertical creates a competitive moat that horizontal marketplaces cannot quickly replicate.

The market opportunity: enterprise legal departments and law firms are significant B2B buyers of AI capabilities — for contract review, legal research, regulatory compliance monitoring, matter management, and M&A due diligence. The AI capabilities available are genuinely varied and technically complex: different AI systems perform meaningfully differently on different legal task types, jurisdictions, and practice areas. Enterprise legal buyers need not just discovery infrastructure but rigorous evaluation that reflects their specific professional requirements.

The trust infrastructure investment: a legal AI marketplace cannot simply adopt the generic security certification framework that applies to enterprise software generally. Legal AI evaluation requires: legal reasoning benchmarks developed and maintained in

partnership with law school AI programs, bar association technology committees, and legal profession standards bodies; jurisdiction-specific performance attestation (a legal AI that performs well on New York contract law may perform differently on California employment law or EU commercial law); bar compliance assessment (which uses of legal AI are permitted under professional responsibility rules in each jurisdiction); and malpractice insurance implications documentation (the legal AI's error rate and the scenarios in which errors are most likely are material to a law firm's professional liability exposure).

The B2B commercial structure for legal AI: law firms and legal departments purchase legal AI capabilities through B2B commercial agreements that differ from standard enterprise software procurement. Law firm professional responsibility rules create specific requirements around client data confidentiality, work product protection, and fee sharing arrangements that must be addressed in the AI vendor's commercial terms. The legal AI marketplace that maintains standard commercial terms updated by legal technology counsel and reviewed against current bar association ethics opinions creates a procurement shortcut that individual law firms would otherwise need to develop independently for each AI vendor relationship.

The retail B2B channel for legal AI: within the legal sector, the retail B2B customer is the small and mid-size law firm or the solo practitioner operating a professional services business — B2B customers by any definition, but with procurement processes that are simpler than large law firms and purchasing behaviors that are more similar to the self-serve developer marketplace customer than to the enterprise IT-led procurement customer. The legal AI marketplace must serve both the Am Law 100 firm with 1,500 attorneys and a sophisticated IT procurement organization, and the 15-attorney firm whose managing partner wants to evaluate a contract review AI with a credit card. Both are B2B customers. Both benefit from the same trust infrastructure. The channel and the commercial structures differ.

CASE STUDY: SALESFORCE APPEXCHANGE

Enterprise Application Marketplace — Trust Infrastructure as the Primary Moat	
The trust architecture	Security Review certification: every application must pass a comprehensive security review before listing. The 'Security Reviewed' badge is a trust signal that enterprise IT security teams rely on as a prerequisite for procurement consideration. This trust infrastructure is what makes AppExchange the de facto procurement channel for Salesforce-integrated applications — not discovery breadth, not price, but the security credential that unlocks enterprise IT approval.
The commercial scale	7,000+ listed applications. $5B+ cumulative transaction volume. 94% of Fortune 500 Salesforce customers use at least one AppExchange application. 25% take rate on app subscription revenue. These numbers represent the commercial outcome of 18 years of marketplace development and trust infrastructure investment — the endpoint of the maturity curve that emerging AI marketplaces are at the beginning of.
The Einstein Trust Layer for AI	Salesforce's AI governance framework for AppExchange AI applications: AI-generated content must be clearly labeled; customer data cannot be used to train foundation models without consent; AI outputs must be traceable to sources. The Einstein Trust Layer is the AI-specific extension of the AppExchange trust architecture — and it is increasingly becoming a listing requirement for AI applications on AppExchange.
The ecosystem flywheel	3.7 million Salesforce-certified professionals who have invested in platform expertise are the network effect that makes AppExchange irreplaceable. Not individual switching costs — ecosystem gravity. The entire professional community of Salesforce administrators, developers, and consultants builds their careers around the AppExchange ecosystem. Moving away from Salesforce means moving away from this professional community, which is a cost that goes far beyond the switching cost of migrating data.

CLOSING

Discovery Is Destiny

The marketplace that surfaces the right capability at the right moment owns the commercial gravity of the AI economy.

Framework F24 — The AI Marketplace Architecture Model

Framework F24 — The B2B AI Marketplace Architecture Model — defines the complete commercial architecture of a B2B AI marketplace platform.

The governing principle: trust enables B2B commerce, and B2B commerce compounds trust. Every platform investment decision should be evaluated against this principle — does it build trust that enables more B2B commercial transactions, and do those transactions generate the resources to build more trust?

The five architecture layers:

Layer 1 — B2B Discovery Infrastructure: the taxonomy (business function, workflow type, integration target, compliance posture), search (bridging the vocabulary gap between enterprise buyer language and AI capability provider language), recommendation systems (surfacing capabilities the enterprise buyer did not know to search for), and evaluation tools (benchmark comparisons, procurement document previews, integration compatibility checks). B2B discovery infrastructure is designed for business buyers making business decisions — not for developers evaluating technical implementations.

Layer 2 — B2B Trust and Compliance Infrastructure: pre-listing certification (security certification verification, compliance documentation, performance attestation), procurement documentation packages (VSQs, DPA templates, standard commercial terms), domain performance attestation (industry-specific benchmarks in partnership with sector organizations), and runtime monitoring with enterprise alerting. The trust infrastructure is the primary competitive moat and the prerequisite for enterprise B2B buyer adoption.

Layer 3 — B2B Transaction and Commercial Infrastructure: retail B2B subscription management (standard commercial terms, automated billing, usage metering), enterprise B2B private offer infrastructure (custom commercial terms, PO-based billing, contract management), B2B agent commerce infrastructure (organizational account governance, spending authority enforcement, micropayment settlement), and take-rate management (tiered rates reflecting value provided at each commercial tier).

Layer 4 — B2B Integration and Composition Infrastructure: enterprise system integration templates (SAP, Salesforce, ServiceNow, Workday, Microsoft 365

connectors), MCP server registry for agent-driven capability discovery, API documentation and SDK provision, and composition testing tools for multi-capability workflow validation.

Layer 5 — B2B Analytics and Intelligence Infrastructure: enterprise buyer usage analytics (what capabilities are being purchased, for what workflows, with what performance outcomes), provider market intelligence (which industries are growing, what capability gaps exist, which certification requirements are emerging), agent commerce analytics (B2B agent transaction patterns, spending attribution, ROI analysis), and marketplace health metrics (certification coverage, trust signal coverage, procurement cycle times).

The three B2B participant types and their requirements:

B2B capability providers need: distribution to enterprise buyers they cannot efficiently reach directly, trust certification that unlocks enterprise procurement approval, and commercial infrastructure (billing, contract management, compliance documentation) that they would otherwise need to build independently.

B2B buyers (retail and enterprise) need: discovery that surfaces relevant capabilities efficiently for their specific business context, trust evidence that enables procurement approval without extended per-vendor security review, and commercial infrastructure that fits their organizational procurement processes.

The marketplace operator needs: take rates and service fees that sustain the trust and commercial infrastructure investment, network effects that compound with scale, and incentive structures that retain both providers and buyers as the platform grows.

Framework F24 — The B2B AI Marketplace Architecture Model — The AI Marketplace Architecture Model

Layer	Function	Investment priority	Success metric	Compounds how
1. Discovery Infrastructure	Taxonomy, search, recommendations,	High — the foundation of	Time to relevant result;	Data flywheel: each search and

	evaluation systems	marketplace value	percentage of searches resulting in deployment	download improves future discovery for all participants
2. Trust and Compliance Infrastructure	Model cards, benchmarks, certifications, runtime monitoring, dispute resolution	Highest — the primary competitive moat for enterprise markets	Percentage of listed capabilities meeting enterprise deployment standard	Trust network: each certified listing makes all listings more trustworthy; each enforcement action validates the trust infrastructure
3. Transaction and Commercial Infrastructure	Pricing APIs, payment processing, take-rate management, A2A settlement	High — enables commercial transactions to occur; required for revenue generation	Transaction success rate; settlement time; dispute rate	None direct — but enables the transactions that generate the data for layers 1 and 2
4. Integration and Composition Infrastructure	SDKs, APIs, MCP servers, integration templates	Medium — reduces friction for providers and buyers; accelerates supply and demand growth	Integration adoption rate; time from listing to first deployment	Integration network: each additional integration creates switching costs that compound the embedding depth
5. Analytics and Intelligence Infrastructure	Usage data, performance monitoring, market intelligence, provider analytics	Medium — becomes more valuable with transaction volume; invest continuously not upfront	Data product revenue; provider engagement with analytics	Data flywheel: more transactions → richer intelligence → better recommendations → more transactions

The B2B AI marketplace is the commercial infrastructure that makes the rest of the B2B AI economy function at scale. Without it, enterprise buyers face an evaluation and procurement burden that slows AI deployment to the pace of their slowest-moving procurement process. Without it, specialized AI capability vendors cannot reach the enterprise buyers who would pay the most for their capabilities without building

commercial organizations that cost more than their initial market can support. Without it, B2B AI agents cannot discover and transact with specialized capabilities at machine speed.

The three archetypes in this book — the developer and ML practitioner hub, the enterprise application marketplace, and the cloud AI service marketplace — each solve a specific version of the B2B discovery, trust, and transaction problem. Each has demonstrated commercial viability. Each has network effects that make them difficult to displace within their specific B2B participant segment.

The specific commercial clarification this book has been deliberate about: B2B marketplace commerce includes a retail channel — the online-originated, self-serve B2B buyer — that is just as B2B in nature as the enterprise direct channel. The developer team that signs up for Hugging Face Pro through the website is a B2B customer at a business organization. The engineering team that subscribes to an AI capability through AWS Marketplace's standard interface is a B2B customer making a commercial decision on behalf of their organization. The retail/enterprise distinction is a channel and complexity distinction within the B2B market — not a business-versus-consumer distinction.

The agent-to-agent commerce layer that is being built right now — in software development, research, and data analytics verticals — will eventually generate more B2B marketplace transaction volume than human-initiated B2B marketplace transactions. Every agent transaction is a B2B transaction, executed by an AI agent on behalf of a business organization, under commercial authority delegated by that organization's governance framework. The B2B marketplace infrastructure required to govern these transactions — organizational account enforcement, spending authority governance, vendor relationship verification, data handling compliance, attribution and accountability — is the infrastructure investment that the most strategically positioned B2B AI marketplace operators are building today.

Build the trust infrastructure first. It is the prerequisite for everything else.

Serve both the retail B2B channel and the enterprise B2B channel from the same platform — they are the same market at different maturity and procurement complexity stages.

Build the agent commerce infrastructure before it is urgently needed — the organizations that establish B2B agent commerce standards will set the commercial norms for a transaction volume that will eventually dominate the marketplace.

In the B2B AI economy, discovery is destiny. The marketplace that surfaces the right capability — to the right enterprise buyer or authorized AI agent, with sufficient trust evidence to act on the discovery, at a price that reflects the value delivered — owns the commercial gravity of the entire ecosystem.

Build the infrastructure. Enforce the trust. Serve the B2B buyer in every channel. That is the marketplace opportunity.

> ***"In the AI economy, discovery is destiny. The marketplace that surfaces the right capability at the right moment owns the commercial gravity of the entire ecosystem."***

THE AI ECONOMY MONETIZATION SERIES · Book Four-B

Framework F24 · The AI Marketplace Architecture Model · Infrastructure Pair with Book Four