

India Foundation Models and the Economics of Multilingual Distribution and Compute

A Decision Framework for Language Coverage Workflow Adoption Infrastructure Cost and Capital

Authored by Chennakeshav Adya

Independent Researcher

September 2026

Abstract

India offers foundation-model companies a large and linguistically diverse distribution opportunity. The commercial case still depends on accepted workflows, lawful data use, reliable deployment and unit economics. Language count, parameter scale, public support and benchmark results provide technical context. They do not establish paid adoption, transferable rights, durable contribution or enterprise value.

This paper develops a decision framework for investors, boards, strategic buyers and public sponsors assessing Indian foundation-model companies. It connects six evidence systems: language and task evaluation, data rights, customer adoption, distribution control, compute economics and capital. The framework treats multilingual capability as a portfolio of language, script, speech, document and domain requirements. It also distinguishes model ownership from the wider service needed to win, deploy and retain customer workflows.

India's public AI infrastructure is expanding. The IndiaAI Mission was approved with an outlay of INR 10,371.92 crore and a public compute objective of at least 10,000 graphics processing units. A March 2026 government release reported that more than 38,000 GPUs had been onboarded. Public access can reduce an important constraint, while productive capacity still depends on allocation, software, data pipelines, interconnect, support and sustained availability [7-13].

Four hypothetical operating cases illustrate the method: a multilingual enterprise application programming interface platform, a regulated-workflow specialist, a public-infrastructure integrator and a voice-and-edge network. The assumptions produce annual revenue from INR 720 crore to INR 1,850 crore, contribution from INR 190 crore to INR 760 crore and cash before financing from negative INR 165 crore to positive INR 95 crore. A separate probability-weighted illustration produces enterprise value of INR 7,540 crore. These figures explain the method. They are not market observations, forecasts or valuation conclusions.

The analysis finds that multilingual distribution creates value when a company can prove language-specific utility, convert access into contracted use, control delivery cost and finance continuing model renewal. Investment and acquisition structures should release capital and consideration against reproducible evaluation, production acceptance, customer commitments, compute availability and collected cash.

JEL Classification: G24, G32, L86, L96, O32, O33

Keywords: India foundation models, Indic languages, multilingual artificial intelligence, sovereign AI, compute economics, workflow adoption, foundation model valuation, IndiaAI Mission

Introduction

India's foundation-model opportunity combines twenty-two scheduled languages, several scripts, extensive code switching, large voice populations and highly varied digital workflows. A model that performs well in Hindi or English may still fail in a Tamil public-service exchange, a Marathi agricultural advisory task, a Bengali financial document, a Telugu customer call or a multilingual engineering record. Commercial diligence needs a defined user, language, task and consequence before it needs a model ranking.

Public infrastructure has expanded the available base. The IndiaAI Mission covers compute, datasets, indigenous models, application development, skills, startup financing and safe and trusted AI [7-13]. AIKosh provides a national platform for datasets, models and development resources [14-16]. AI4Bharat has released open datasets, translation systems and evaluation resources for Indian languages [17-22]. These initiatives can lower development barriers. Company value still requires evidence that a specific team converts those resources into accepted work and cash.

The provider landscape includes different technical and commercial positions. Sarvam describes models trained in India and a full-stack sovereign platform [23-27]. Tech Mahindra reports its Project Indus work on Hindi and dialects [28-31]. BharatGen describes multilingual models and public research outputs [32-34]. Provider statements form hypotheses for diligence. Independent tests, contracts, workload telemetry and collections determine whether those capabilities support an investment case.

The decision for an investor or board is whether a company controls a repeatable system for turning multilingual capability into accepted outcomes, contribution and cash. That system includes data rights, evaluation, product integration, distribution, deployment, compute and model renewal. This paper connects those elements to valuation and transaction design.

1 Define the investment decision and perimeter

The first document in the diligence room should state the decision. A growth investor may be deciding whether to fund a model owner, an application company or a sovereign deployment platform. A strategic buyer may seek data, talent, customer access or operating control. A public sponsor may want national capability, Indian-language inclusion, resilience or productivity. These objectives create different value tests.

The perimeter should identify legal entities, shareholders, repositories, model weights, tokenisers, data pipelines, evaluation systems, deployment software, contracts, compute agreements and customer obligations. It should show which assets are owned, licensed, open-weight, partner-controlled or customer-specific. Shared branding and national association do not resolve legal ownership or transferability.

The operating model also matters. A company may train a general model, adapt third-party weights, orchestrate several models, sell an application or operate a private instance. Each position creates different capital needs and defensibility. The valuation should allocate revenue, cost and risk to the actual layer rather than to an undifferentiated Indian-language AI narrative.

2 Segment multilingual capability before measuring performance

Indian-language capability should be segmented by language, script, dialect, register, code switching, transliteration, speech, optical character recognition, handwriting, retrieval and domain terminology. Scheduled-language coverage is a useful inventory. It does not show whether a model performs the task, population and format required by a customer. Each revenue cohort needs a narrower capability definition.

Evaluation should preserve the user population, input quality and consequence of error. IndicGenBench covers generation tasks across twenty-nine Indian languages, while IndicQA and MILU test other aspects

of question answering and culturally relevant knowledge [18-22]. Published research reports meaningful gaps across languages and tasks. A composite average can hide a severe weakness in a commercially important language or workflow.

The customer test uses the proposed model version, prompt, retrieval, tools, latency, security and human-review rules. It should include code switching, named entities, numerals, dates and local terminology where they affect performance. Results should be retained by cohort so that later model changes can be compared with the accepted baseline.

3 Build the corpus rights ledger

Training and adaptation data should be recorded by source, owner, licence, purpose, jurisdiction, collection method, permitted use, retention, deletion and model version. Public availability is not a substitute for permission. Contracts should cover model development, commercial use, derivatives, sublicensing, transfer in a transaction and audit where relevant.

Indic-language corpora can contain books, news, public records, social media, transcribed speech, call-centre recordings, educational material, technical manuals and customer documents. Each class carries different rights and privacy questions. Dialect and speech datasets may contain personal data or sensitive context. Customer-provided data may support retrieval or fine-tuning without creating a transferable company asset.

The ledger should distinguish raw data, cleaned datasets, annotations, synthetic data, feedback and derived evaluation sets. It should identify duplicate or contaminated benchmark material. A buyer should be able to trace a released model to the relevant dataset versions and approvals. Where provenance is incomplete, the model should include the cost and time needed to remove, replace, retrain and re-evaluate the affected capability.

4 Value proprietary corpora through replacement and revenue evidence

A dataset has strategic value when lawful access is scarce, the data improves a required capability and the company can use that improvement commercially. Volume alone is a weak measure. A smaller, current and carefully annotated corpus can be more useful than a larger body of duplicated or low-quality text. Replacement cost should include sourcing, rights, cleaning, annotation, quality assurance, storage, evaluation and elapsed time.

Income evidence requires a link from corpus to workflow. The company should show that a defined dataset improved acceptance, reduced review effort, increased conversion, supported a higher price or protected retention. The analysis should control for model, prompt, retrieval and product changes. Customer-specific improvements may be valuable to the contract while remaining unavailable for the broader platform.

The valuation should avoid counting the same advantage twice. Corpus benefits can appear in forecast revenue, contribution or option value. An additional asset premium is appropriate only when the method excludes those cash flows or recognises a distinct transferable right. IAS 38 and IFRS 13 provide relevant accounting and fair-value principles, while transaction analysis still requires company-specific legal and financial advice [1-5].

5 Design a dialect and domain evaluation system

The evaluation system should start with use-case acceptance criteria. For a regulated document workflow, the measures may include extraction accuracy, grounded answers, citation completeness, prohibited disclosure, latency, cost and human-review time. For speech, word error rate is useful, although task

completion and escalation may be more important to the customer. For content generation, factuality, terminology, tone and approval effort can determine value.

Test populations should represent countries, age groups, accents, code switching, script variation and document quality where these factors affect use. Native reviewers need clear rubrics and adjudication procedures. Inter-reviewer agreement should be measured. Customer confidential examples require access controls and a permitted-use record.

The company should maintain versioned scorecards and failure taxonomies. Improvement on an average benchmark should not conceal regression in a high-value cohort. Release decisions should include capability, safety, cost and contractual commitments. Evaluation becomes an operating asset when it shortens customer diligence and guides model routing, adaptation and human review.

6 Convert model benchmarks into distribution evidence

The evidence ladder begins with reproducible evaluation and progresses through demonstration, controlled pilot, user acceptance, production, invoice, collection and renewal. Each stage supports a different conclusion. Downloads, announced partnerships, memoranda and subsidised experiments may demonstrate interest or learning. Contracted production and collected cash support the commercial case.

Distribution should be mapped by route: direct enterprise sales, government programmes, states, cloud marketplaces, telecommunications providers, systems integrators, software partners and developer adoption. The ledger should state which party controls the customer relationship, data, billing, implementation, support and renewal. A distribution agreement has limited value when minimum activity, economics and account ownership are unclear.

The adoption file should record customer entity, sector, state, language, workflow, model version, acceptance threshold, users, volume, price, direct cost, implementation effort and cash. Forecast conversion should follow observed cohorts. Free or grant-supported use should remain distinct from commercial revenue while its technical learning is recorded.

7 Measure customer adoption and retention

Usage should be measured at the level that customers buy. Relevant units can include accepted documents, resolved contacts, completed applications, analyst hours saved or transactions reviewed. Token volume may reconcile compute cost, while it rarely explains customer value on its own. Free usage and promotional credits should be separated from paid activity.

Retention analysis should use contracted cohorts and distinguish model, application and service revenue. A customer may renew because of data integration, support or switching cost even when the underlying model is substitutable. Conversely, an open-weight deployment may preserve customer control while reducing recurring licence revenue. The company should explain the source of retention and the cost of maintaining it.

Pricing evidence should include list price, discount, implementation, minimum commitment, capacity reservation, support, acceptance credits and collection. Revenue recognition and cash timing can differ materially. IFRS 15 provides a framework for identifying contracts and performance obligations; company-specific accounting conclusions require professional judgement [4].

8 Define sovereign and public infrastructure option value

Sovereign option value should be decomposed into rights and operating capabilities. Relevant elements include Indian legal control, model and software rights, data location, key control, compute access, release authority, security operations, continuity, portability and the ability to support public or regulated customers during external disruption. Hosting in India supplies one component of that position.

Public infrastructure can create option value through affordable compute, data access, procurement pathways and shared language resources. The IndiaAI Compute Portal lists cloud, storage, model operations and language services for eligible users [10-13]. The valuation should reflect the actual allocation, subsidy period, service terms, productive utilisation and renewal risk. An announced national capacity number is not the same as capacity contractually available to one company.

The customer's requirement determines economic value. A government workflow may require local administration, audit and continuity. A regulated institution may require privacy, explainability, human review and tested recovery. A commercial workflow may prioritise language quality, latency and total cost. Each claimed sovereign advantage should be tied to eligibility, revenue, cost or a measurable reduction in risk.

9 Test private deployment and portability

Private deployment can take the form of customer cloud, sovereign cloud, dedicated hosted capacity, on-premise infrastructure or a hybrid system. The diligence team should reproduce a representative workload in the proposed environment. It should measure quality, latency, throughput, security, monitoring, recovery, update process and cost.

The portability test should identify model licences, container images, proprietary dependencies, tokenisers, retrieval services, moderation, observability, accelerators and network requirements. Control of weights does not establish control of the complete service. The customer and supplier should agree who operates each layer and who can restore it after an incident.

Financial analysis should separate licence, implementation, capacity, managed service and support. Private deployments can create larger contracts and longer retention. They can also require bespoke engineering, inventory, parallel model versions and working capital before acceptance. Minimum prices and maintenance reserves should reflect the actual configuration.

10 Reconstruct compute and inference economics

Compute analysis should reconcile public allocations, commercial cloud, reserved accelerators and owned infrastructure. Evidence includes allocation letters, provider invoices, accelerator type, effective availability, utilisation, storage, networking, software, support and termination rights. The IndiaAI ready reckoner illustrates that different workloads and configurations carry different hourly prices and operating characteristics [11]. A headline GPU count does not establish productive capacity.

Inference cost should be measured per accepted workflow unit. The schedule should include input and output tokens, speech, retrieval, tools, caching, guardrails, retries, human review, idle capacity and service credits. Language mix can change tokenisation and task length. Voice and document workflows add preprocessing and post-processing that a token-only cost measure can miss.

Training economics require a programme ledger. Each run should record objective, dataset versions, compute, elapsed time, evaluation, release decision, customer relevance and cash effect. Public subsidy should be shown separately from underlying resource cost. The capital plan should identify the funding required to reach the next technical and commercial threshold if access terms or market prices change.

11 Separate open weight access from owned advantage

Open-weight models can lower development cost, support private deployment and widen the ecosystem. They can also make baseline capability easier to reproduce. The diligence team should read the applicable licence and verify the exact version. Open access to weights does not imply unrestricted rights to every dataset, tool, trademark or hosted service.

A company using open weights can still build durable value through lawful domain data, evaluation, workflow integration, distribution, support and controlled deployment. Each advantage should be tested for transferability and customer relevance. Adaptation artefacts and customer fine-tunes require clear ownership and confidentiality terms.

The substitution test should ask how long a competent competitor would need to reproduce the accepted workflow with another model. It should include data access, evaluation, integration, security approval, customer migration and support. A short technical replacement can still involve a long commercial transition. The valuation should model both.

12 Reconcile privacy security and human accountability

Indian foundation-model workflows can process personal, confidential and regulated information. The Digital Personal Data Protection Act 2023 and the Digital Personal Data Protection Rules 2025 establish the national data-protection framework and phased implementation requirements [35-38]. Duties depend on the entities, processing purpose, data and commencement provisions. Specific deployments require current legal advice.

Controls should cover purpose, notice, consent where applicable, minimisation, access, encryption, retention, erasure, transfer, incident response and supplier management. Training, retrieval, logging and feedback are distinct processing activities. Customer prompts and outputs should not enter training or evaluation datasets without an authorised basis, contractual clarity and technical enforcement.

The Reserve Bank of India's FREE-AI work addresses responsible and ethical AI in the financial sector [39-41]. NIST and ISO provide complementary risk-management structures [42-47]. Human accountability remains necessary for consequential decisions. A company should define review, escalation, override and incident ownership and measure whether controls operate in each deployed language.

13 Assess distribution power and public procurement

India's routes to market differ by customer and state. They can include direct enterprise contracting, central and state government procurement, public digital platforms, cloud marketplaces, telecommunications providers, banks, systems integrators and software companies. The forecast should show who controls the account, billing, implementation, data and renewal in each route.

Public programmes can provide access, reference deployments and shared infrastructure. They can also introduce eligibility conditions, milestone acceptance, budget timing and concentration. A memorandum or approved project should be classified by its enforceable scope, funding, acceptance process and payment mechanism. Government association alone does not establish recurring revenue.

Distribution value should follow observed conversion, contribution and retention. Revenue sharing, implementation cost, account ownership, exclusivity and termination rights belong in the cohort model. The downside case should include delayed awards, policy changes, reduced subsidy, channel margin pressure and partner substitution.

14 Value talent and the multilingual research operating system

Indian-language foundation-model development depends on researchers, data engineers, linguists, evaluators, safety specialists, infrastructure engineers and product teams. Critical knowledge can sit with a small group. The company should map roles, employment and invention rights, access privileges, succession, retention and the time needed to replace each capability.

The research operating system should preserve experiment records, data versions, code, model artefacts, evaluation, release decisions and incident learning. A buyer should be able to reproduce a defined build

and explain changes between releases. Documentation that exists only in personal accounts or informal messages weakens transferability.

Talent value should be reflected through the cost of retaining and rebuilding capability, not through a separate premium that ignores future compensation. The forecast should include competitive pay, visas where relevant, compute support and management capacity. A strong system reduces key-person exposure and converts individual expertise into organisational capability.

15 Construct four India foundation model cases

The multilingual enterprise application programming interface platform competes on coverage, developer access, reliability and price. The regulated-workflow specialist serves a smaller number of banking, insurance, healthcare or public-administration tasks with stronger evaluation and documentation. The public-infrastructure integrator operates controlled model services for government and regulated customers. The voice-and-edge network focuses on speech, assisted access and lower-bandwidth distribution.

Each case should be modelled by customer cohort. Assumptions include conversion, accepted volume, price, compute, implementation, support, research, data, privacy, working capital and financing. Language expansion increases addressable workflows while adding evaluation, support and release obligations. Distribution can reduce customer-acquisition cost while giving a partner control over economics and renewal.

The cases are decision tools. Management should replace every assumption with company evidence and assign an owner. The board should define events that move the plan between cases, including language-specific acceptance, production conversion, signed minimum commitments, compute allocation, channel performance or delayed collections.

16 Build the integrated financial model

The model should connect technical and commercial ledgers to revenue, contribution, operating expense, capital expenditure, working capital and financing. Customer revenue should reconcile to contract, acceptance, invoice and cash. Compute should reconcile to usage and provider terms. Research and corpus spending should reconcile to the programme plan.

Cash timing can determine feasibility. A public or regulated deployment may require evaluation, security, integration and dedicated capacity before customer acceptance. The model should show milestone billing, receivables, supplier terms and liquidity by programme. A profitable contract can create a financing gap when cost precedes collection.

The base case should use observed cohort performance. Upside should remain conditional on evidence and funding. Downside should combine weaker conversion, lower price, higher inference and support cost, rights remediation, capacity delay and a slower financing round. Management responses should be specific and costed.

17 Apply valuation methods to the evidence states

Discounted cash flow can value established cohorts when revenue, cost and reinvestment are supportable. Market approaches can provide context, although differences in model ownership, data rights, deployment, growth, margin and capital intensity limit simple revenue multiples. A replacement-cost analysis can help assess corpora and engineering assets without proving income value.

The option method is useful when value depends on future technical and commercial gates. States can include accepted pilots, repeatable paid deployments, scaled cohorts and strategic distribution. Each state

needs a probability, timing, capital requirement and enterprise value based on its economics. Probabilities should change when observable events occur.

The methods should reconcile. The valuation date, currency, debt, leases, commitments, grants, credits and non-operating assets should be explicit. Public support or compute access should be included only under its terms and available duration. IFRS 13 and International Valuation Standards provide relevant principles for fair-value and valuation work [1-3].

18 Illustrative India foundation model economics

The hypothetical enterprise platform has annual revenue of INR 1,850 crore and contribution of INR 760 crore. Its scale supports the largest research, data and compliance programme, assumed at INR 620 crore. The regulated specialist records INR 1,250 crore of revenue and INR 475 crore of contribution. Documentation, integration and specialist support keep cash before financing at negative INR 40 crore.

The public-infrastructure integrator records INR 980 crore of revenue, INR 310 crore of contribution and negative INR 120 crore of cash before financing. The voice-and-edge network records INR 720 crore of revenue, INR 190 crore of contribution and negative INR 165 crore of cash. Lower revenue does not by itself imply lower strategic value. The investor must assess whether language access, distribution and future cohorts justify the capital required.

Every number is a management assumption created for illustration. The cases are not observations about any provider and do not predict the Indian market. A real analysis requires contracts, acceptance records, workload telemetry, compute terms, payroll, data rights, invoices, collections and a funded operating plan.

19 Value data rights adoption and sovereign options separately

The framework separates three value sources. Data rights support reproducible capability and remediation control. Adoption evidence supports revenue quality and customer utility. Sovereign options support access, continuity and strategic flexibility for defined buyers. The same fact should not be credited in more than one category without a distinct economic effect.

A rights score should measure chain of title, permitted use, transferability, traceability and remediation. An adoption score should measure acceptance, production, contribution, collection and renewal. A sovereign-options score should measure control, portability, capacity, security and customer eligibility. Weightings depend on the investment thesis.

The scores are governance aids rather than valuation formulas. They help identify missing evidence and conditions. Cash-flow and option models translate verified effects into value. The committee should document which evidence changes revenue, cost, risk, probability or capital.

20 Structure investment and consideration around evidence

Investment tranches can align funding with corpus clearance, reproducible evaluation, accepted deployment, minimum customer commitments and capacity. Acquisition consideration can combine base value, escrow, holdback and contingent payments. Milestones should use objective data sources and defined measurement periods.

Representations should address ownership, licences, data provenance, privacy, model versions, security, customers, compute and employment rights. Indemnities allocate financial exposure while leaving operational remediation to the company. A rights exception that threatens a core capability may require a closing condition or price adjustment rather than a general warranty.

The capital plan should protect customers if a later tranche is not released. Boards should avoid structures that create an unfunded service obligation. Instruments, tax and accounting treatment require current professional advice in the relevant jurisdictions.

21 Govern value after investment

Post-investment governance should use the same evidence architecture as diligence. The board pack should report corpus-rights exceptions, language and domain evaluation, accepted workflows, contracted cohorts, contribution, collections, compute, incidents, runway and financing. Model releases and material customer configurations should have accountable approval.

The company should refresh the valuation case after material changes in model, data, deployment, customer use, regulation, capacity or funding. Historical scores can become stale after a new release or supplier change. A dated exception register should show owner, remediation and effect on customer commitments.

Governance should remain decision oriented. The board needs the variables that change capital allocation, customer obligations and enterprise value. Detailed ledgers provide auditability. The principal report should explain the economic consequence and the required action.

Conclusion

Indian multilingual foundation-model value rests on a connected operating system. Lawful and transferable data supports capability. Task-specific evaluation demonstrates fitness. Customer acceptance, invoice, collection and renewal demonstrate adoption. Tested deployment, security and portability support control. Compute and funding determine whether the system can continue.

Benchmarks, national association and parameter scale remain useful signals. Their economic meaning depends on the rights, workflows and cash they support. An investor should value verified customer cohorts and controlled options while funding technical progress through explicit evidence gates.

The proposed framework gives boards a practical way to separate corpus value, adoption value and sovereign option value. It also keeps hypothetical scenarios distinct from observed evidence. The resulting transaction case can be updated as rights are cleared, deployments are accepted and cash is collected.

Indian multilingual model rights and provenance register

The register should contain source, owner, licence, jurisdiction, collection method, personal-data class, permitted training and adaptation, commercial-use rights, sublicensing, transfer, retention, deletion, dataset version, model version and approval. It should include raw, cleaned, annotated, synthetic and feedback data. Every exception should have an owner, customer effect and remediation plan.

The company should test the register by tracing a selected model release back to dataset evidence. A static policy is insufficient when source records cannot be reproduced. The test should also cover data added after the principal training run and customer-specific fine-tunes. Access and confidentiality should reflect contractual duties.

For material gaps, management should estimate replacement data, labelling, compute, evaluation, delay and customer consequences. The schedule should distinguish a removable source from a dataset central to Indian-language or domain capability. That distinction affects warranties, reserves, price and closing conditions.

Language domain and workflow evaluation file

The file should define user population, country, dialect, register, script, domain, task, consequence, model, prompt, retrieval, tools, human review, acceptance threshold and test date. It should retain examples, reviewer qualifications, agreement, failures and release decisions. Benchmark contamination and overlap with training data should be checked.

Customer tests should remain distinct from public benchmarks. Public suites help compare technical capability under published conditions. Customer acceptance measures performance inside the contracted workflow. Both are useful when their scope and limitations are stated.

The file should connect each accepted evaluation to product configuration, contract and support obligation. A later release should be compared with the accepted baseline. Material regression should trigger a release decision and customer communication under the applicable contract.

Adoption contract and collection ledger

The ledger should trace each customer from qualified opportunity through diligence, pilot, acceptance, production, invoice, collection and renewal. It should contain entity, sector, country, workflow, model, language variant, volume, price, discount, direct cost, implementation effort, partner share, receivable and accountable owner.

Forecast accuracy should be measured by cohort. Management should compare original probability, award date, production date, revenue, contribution and collection with actual results. Persistent delay should change pipeline conversion and working-capital assumptions.

Free, grant-funded, strategic and commercial workloads should be identified separately. A high-usage free service can create technical learning without proving pricing. A government programme can create procurement access while depending on budget and acceptance. The board should see the economic status of each cohort.

Distribution deployment and continuity dossier

The dossier should state legal ownership, hosting, model and software rights, administrator access, cryptographic key control, compute priority, data location, subcontractors, support, incident response, recovery, portability and exit. Each assertion should link to a contract, test or controlled operating record.

A representative workload should be restored in the proposed environment and timed. The exercise should cover model artefacts, retrieval, moderation, monitoring, identity, secrets, network and data. Exceptions should show the customer and service consequence.

The dossier should distinguish control from location. Infrastructure in the India can remain dependent on a foreign supplier for capacity, software or administration. A distributed component can still be governed through enforceable rights and tested alternatives. The commercial claim should match the actual control position.

Compute capital and transaction file

The compute file should reconcile reservations, available accelerators, utilisation, invoices, credits, storage, networking, energy, support, location, service level, renewal and termination. Forecast capacity should match customer volumes and research programmes. Constrained and downside cases should include delayed procurement and lower utilisation.

The capital file should show programme funding, working capital, debt and equity by evidence gate. It should identify obligations that survive if a grant, allocation or financing round is delayed. The board should preserve sufficient liquidity for contracted customer service and orderly remediation.

The transaction room should connect title, data rights, model versions, evaluation, customers, compute, security, employees, financing and forecast vintages. Every committee claim should link to a dated source and accountable owner. This linkage allows the valuation case to change when the evidence changes.

The financial model should also preserve a bridge between technical events and cash. A new model release can increase research cost before it changes customer revenue. A rights exception can require retraining, evaluation and customer communication. A capacity reservation can improve continuity while creating a fixed commitment. Each event should have a dated assumption, an accountable owner and a defined line in the forecast. This bridge lets the board distinguish a technical milestone from a financially funded operating plan.

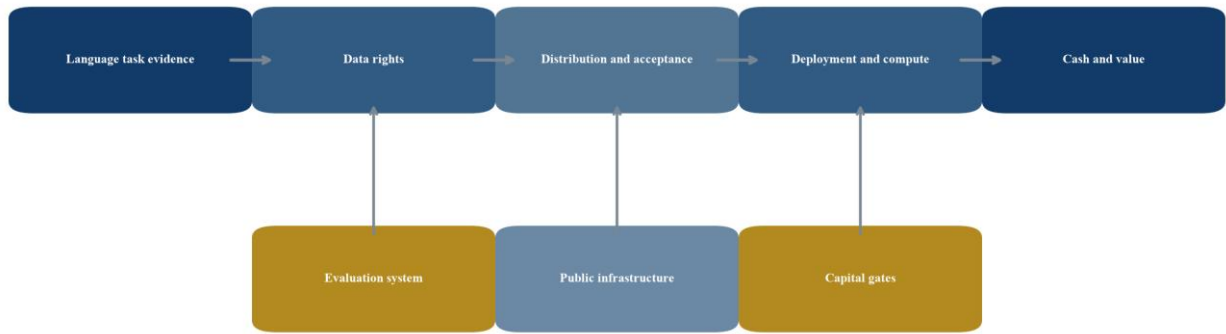
Management should maintain a forecast-vintage file. The file should record the customer, workload, volume, price, direct cost, implementation, collection and financing assumptions used at each decision date. Actual results should be compared with the original case and the latest revision. Persistent optimism in conversion, delivery or collection should reduce the weight given to management estimates. Forecast discipline matters because a model company can consume substantial cash before the difference between technical interest and paid adoption becomes visible.

The transaction file should include a severe but plausible continuity case. The case can combine loss of a corpus licence, delay in a major customer, temporary capacity constraint, required security remediation and slower financing. The company should identify the services it would preserve, the research it would defer, the cash needed for customers and the decisions available to the board. This exercise tests whether sovereign positioning is supported by operational resilience and whether the proposed capital structure can carry the obligations created by customer contracts.

The committee should close the evidence loop after investment. Conditions, indemnities, holdbacks and milestones should be assigned to owners and reviewed against the same records used in diligence. If a customer acceptance gate releases consideration, the acceptance evidence should reconcile to the product version, contract and cash schedule. If a rights remediation closes an exception, the company should show the affected data, model and evaluation path. Governance that retains this traceability makes later financing, acquisition and regulatory review more reliable.

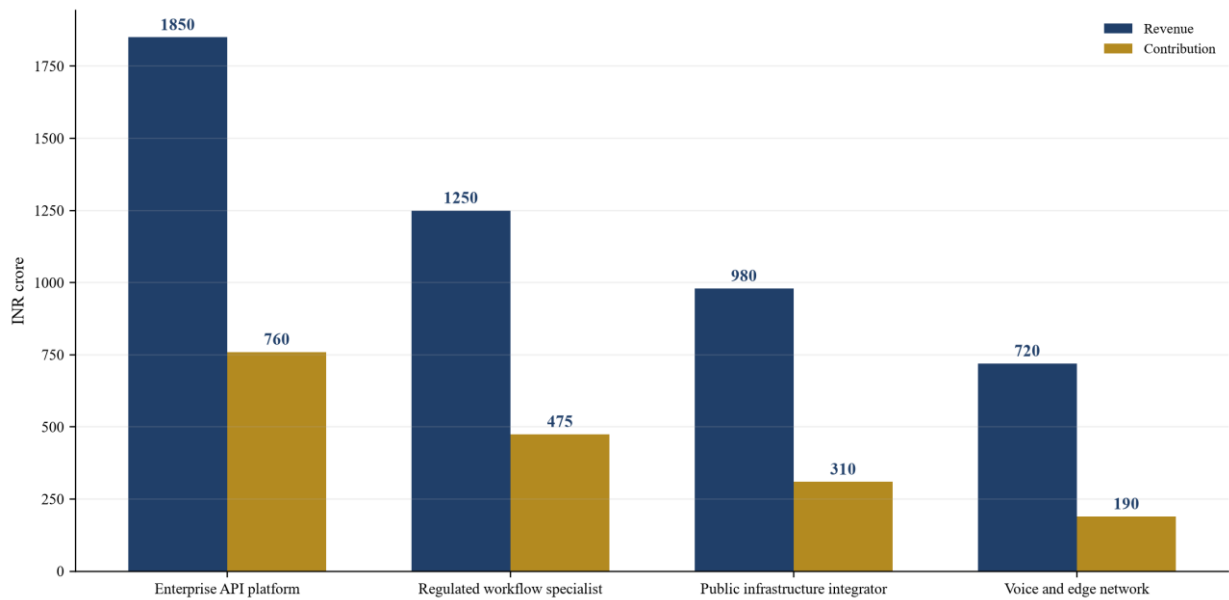
Finally, the evidence room should separate observations from scenarios. Contracts, invoices, collections, model artefacts, licences and test results are observations. Conversion rates, renewal, future price, research success and enterprise value remain assumptions until events occur. Each scenario should state its basis, date and sensitivity. This separation helps the investment committee update probabilities without rewriting the historical record or presenting a hypothetical outcome as a current market fact.

Figure 1 India foundation model value architecture



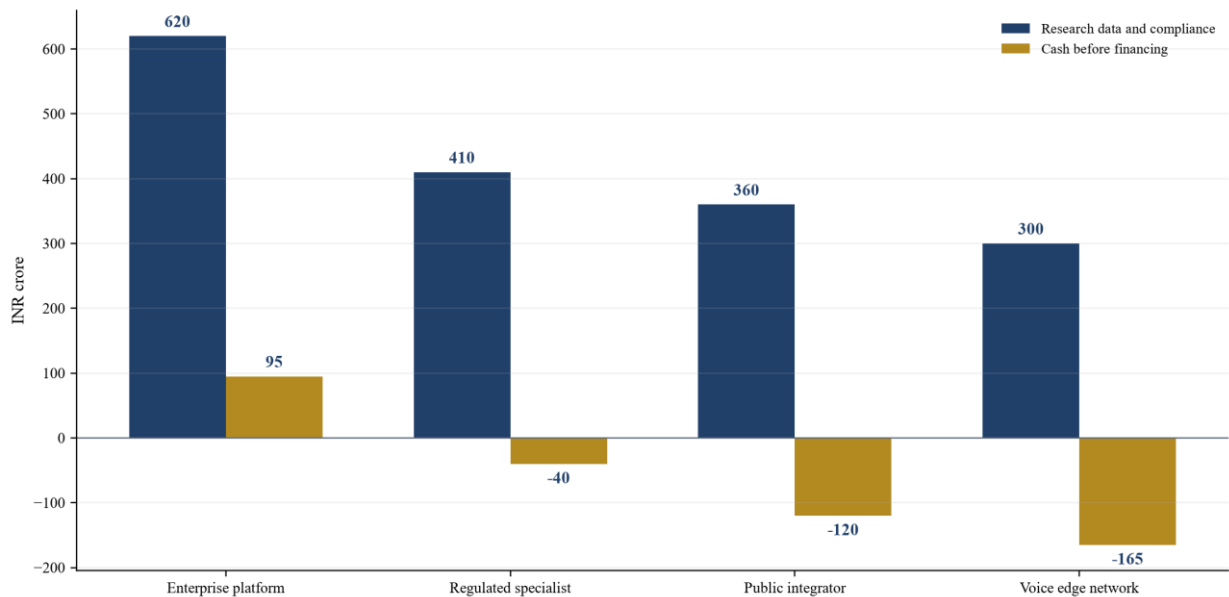
Proposed framework; conclusions require company-specific technical commercial legal and financial evidence.

Figure 2 Hypothetical annual revenue and contribution by India foundation model case



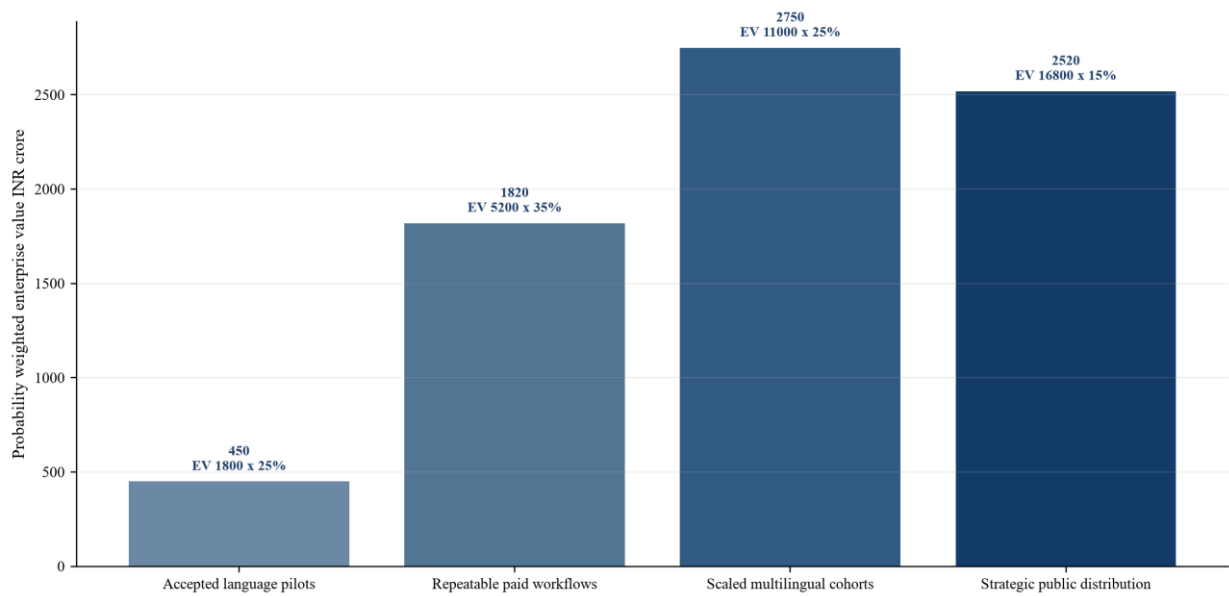
Management assumptions in INR crore; figures are not market observations forecasts or valuation conclusions.

Figure 3 Hypothetical programme cost and cash before financing



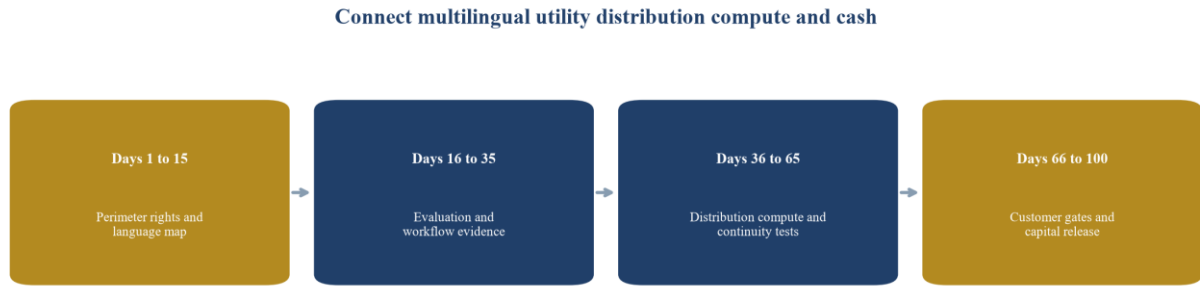
Management assumptions in INR crore; negative cash indicates illustrative financing need.

Figure 4 Hypothetical probability weighted enterprise value by evidence state



Management assumptions in INR crore; total probability weighted value is INR 7540 crore.

Figure 5 First hundred days evidence and capital sequence



Proposed sequence; timing should follow customer regulatory infrastructure and financing constraints.

Table 1 India foundation model investment perimeter

Component	Required evidence	Valuation question	Principal risk
Models and software	repositories weights licences assignments release records	what capability is owned and transferable	open or restricted dependencies
Data and evaluation	provenance licences versions language and task tests	can capability be reproduced and defended	missing rights or unsupported performance
Workflow adoption	acceptance contracts invoices collections renewals	which capabilities create paid demand	pilots mistaken for recurring economics
Distribution and compute	channel terms allocations utilisation hosting continuity	can customers be reached and served economically	partner capacity or subsidy concentration
Governance and security	privacy role map tests incidents approvals	can obligations be evidenced at scale	remediation interruption or liability

Proposed diligence structure; requirements depend on the company customer and transaction.

Table 2 Multilingual capability segmentation

Segment	Example evidence	Commercial use	Diligence limitation
Formal written language	controlled reasoning extraction and generation tests	government media education and documents	average score can hide domain failure
Regional language and speech	native reviewer dialogue and speech evaluation	service centres assistants and field operations	language labels can hide accent variation
Code switching and romanisation	mixed-language and romanised task sets	commerce support and digital channels	spelling and tokenisation vary
Domain language	customer corpus terminology and acceptance	banking law health engineering	confidential tests may not transfer
Document and multimodal	audio document image and task-completion evidence	public services archives and operations	component accuracy differs from workflow utility

Each segment requires a task and customer-specific evidence standard.

Table 3 Distribution and customer economics ledger

Field	Required evidence	Control	Decision use
Validation and acceptance	protocol outputs review sign off	versioned model workload and language	conversion and utility
Contract and billing	signed order price term invoice service level	entity and term reconciliation	revenue and renewal
Channel and delivery cost	partner share compute implementation evaluation support	invoice and allocation reconciliation	contribution and cash
Collection	receivable settlement bank receipt	aged balance and exception owner	liquidity and customer quality
Retention	renewal usage accepted outcome and support	cohort and configuration continuity	durability of workflow value

Proposed minimum evidence fields for customer-level analysis.

Table 4 Hypothetical annual operating cases

Case	Revenue	Contribution	Research data and compliance	Cash before financing
Multilingual enterprise API platform	1850	760	620	95
Regulated workflow specialist	1250	475	410	-40
Public infrastructure integrator	980	310	360	-120
Voice and edge network	720	190	300	-165

Management assumptions in INR crore; figures are not market observations forecasts or valuation conclusions.

Table 5 Hypothetical regulated workflow contribution bridge

Item	INR crore	Evidence required
Revenue	1250	customer contract acceptance invoice and cash ledger
Compute and model delivery	-335	workload telemetry provider invoice and allocation
Data evaluation and compliance	-155	people supplier rights and programme records
Implementation partner and support	-285	project time partner settlement and service evidence
Contribution before central cost	475	reconciled customer cohort schedule

Management assumptions in INR crore; excludes central research sales administration financing and tax.

Table 6 Hypothetical valuation evidence states

Evidence state	Enterprise value	Probability	Weighted value
Language rights mapped and accepted pilots	1800	25%	450
Repeatable paid workflows	5200	35%	1820

Evidence state	Enterprise value	Probability	Weighted value
Scaled multilingual cohorts	11000	25%	2750
Strategic public and enterprise distribution	16800	15%	2520
Total		100%	7540

Management assumptions in INR crore; this is not a valuation conclusion.

Table 7 Investment and transaction evidence gates

Gate	Required evidence	Capital or consideration response	Post decision measure
Rights and control	assignments licences data provenance reproduced build	closing condition or protected tranche	exceptions and controlled release
Workflow adoption	accepted pilot signed order production invoice collection	base value after verification	conversion contribution and cash
Language and domain quality	controlled evaluation and customer acceptance	milestone funding	accepted outcome and review cost
Privacy security and accountability	role map tests incidents recovery and remediation	reserve holdback or staged release	exceptions and closure
Compute and distribution	allocation terms utilisation cost channel and exit test	funding reserve or price adjustment	accepted task cost and continuity

Proposed framework; actual instruments require current legal tax accounting and financial advice.

Frequently asked questions

What creates value in an Indian foundation model company

Value follows lawful and transferable data rights, accepted performance in defined language workflows, controlled distribution, repeatable deployment, durable contribution and sufficient capital to maintain capability. Language count and benchmarks should be connected to customer and cash evidence.

How should multilingual data assets be valued

The analysis should test rights, scarcity, replacement cost, capability improvement and customer economics. A dataset should not receive a separate premium when its benefit is already included fully in forecast cash flow.

Does access to subsidised compute support a valuation premium

It can improve economics and reduce financing needs. The investor should verify allocation, accelerator type, productive availability, subsidy duration, software and network cost, eligibility and renewal before reflecting the benefit.

How should language benchmark results be used

They support a technical hypothesis. A commercial conclusion requires reproducible results on the target language, script, domain and workflow followed by customer acceptance, contribution and retention evidence.

What does sovereign AI mean in India

It should be defined through specific rights and capabilities across ownership, data, model, compute, administration, security, continuity and portability. National hosting supplies one component of that position.

What is the principal downside case

The useful downside combines slower customer conversion, language-specific failure, data remediation, higher inference and support cost, reduced subsidy or capacity, channel concentration, longer collections and constrained financing.

Which evidence should release investment capital

Useful gates include cleared rights, reproducible evaluation, accepted deployment, signed minimum commitments, contracted or allocated compute, invoicing and collected cash. Transaction documents should define measurement and dispute procedures.

Which Matchpoint practice is relevant to this work

This research connects to Matchpoint Partners' strategy and execution, valuation, capital raising, transaction diligence and mergers and acquisitions work for technology companies, investors, public institutions and strategic partners.

References

- [1] IFRS Foundation. IFRS 13 Fair Value Measurement. <https://www.ifrs.org/issued-standards/list-of-standards/ifrs-13-fair-value-measurement/>
- [2] IFRS Foundation. Educational material on measuring unquoted equity instruments within IFRS 13. <https://www.ifrs.org/content/dam/ifrs/supporting-implementation/ifrs-13/education-ifrs-13-eng.pdf>
- [3] International Valuation Standards Council. International Valuation Standards. <https://www.ivsc.org/standards/>
- [4] IFRS Foundation. IFRS 15 Revenue from Contracts with Customers. <https://www.ifrs.org/issued-standards/list-of-standards/ifrs-15-revenue-from-contracts-with-customers/>
- [5] IFRS Foundation. IAS 38 Intangible Assets. <https://www.ifrs.org/issued-standards/list-of-standards/ias-38-intangible-assets/>
- [6] IFRS Foundation. IAS 36 Impairment of Assets. <https://www.ifrs.org/issued-standards/list-of-standards/ias-36-impairment-of-assets/>
- [7] Press Information Bureau Government of India. Cabinet approves IndiaAI Mission. 7 March 2024. <https://www.pib.gov.in/pressreleaseiframepage.aspx?lang=2&prid=2012355®=3>
- [8] Press Information Bureau Government of India. IndiaAI Mission outlay and compute infrastructure. 7 March 2024. <https://www.pib.gov.in/Pressreleaseshare.aspx?PRID=2012375&lang=2®=48>
- [9] Ministry of Electronics and Information Technology. IndiaAI Expert Group Report. <https://www.meity.gov.in/writereaddata/files/IndiaAI-Expert-Group-Report-First-Edition.pdf>
- [10] IndiaAI. IndiaAI Compute Portal. <https://compute.indiaai.gov.in/login>
- [11] IndiaAI. Ready Reckoner for Compute Users. March 2026. <https://compute.indiaai.gov.in/files/Ready-Reckoner.pdf>
- [12] IndiaAI. Compute end-user allocations. <https://www.compute.indiaai.gov.in/endusers>
- [13] Press Information Bureau Government of India. IndiaAI Mission expands AI ecosystem with affordable compute and startup support. 25 March 2026. <https://www.pib.gov.in/PressReleasePage.aspx?PRID=2245069&lang=1®=1>
- [14] AIKosh. About the national datasets platform. <https://aikosh.indiaai.gov.in/home/about-us/>

- [15] AIKosh. IndiaAI Mission and university engagement. <https://aikosh.indiaai.gov.in/workshop>
- [16] IndiaAI Impact Summit. Equitable access to foundational AI resources. https://impact.indiaai.gov.in/assets/INDIA_Themes.pdf
- [17] AI4Bharat IIT Madras. Research programme. <https://ai4bharat.iitm.ac.in/>
- [18] AI4Bharat IIT Madras. Machine translation research and IndicTrans2. <https://ai4bharat.iitm.ac.in/areas/nmt/>
- [19] Singh H and others. IndicGenBench. 2024. <https://arxiv.org/abs/2404.16816>
- [20] Singh A K and others. IndicQA Benchmark. 2024. <https://arxiv.org/abs/2407.13522>
- [21] Verma S and others. MILU multilingual understanding benchmark. 2024. <https://arxiv.org/abs/2411.02538>
- [22] Aravapalli A and others. IndicSentEval. 2024. <https://arxiv.org/abs/2410.02611>
- [23] Sarvam AI. Sarvam M technical report. 23 May 2025. <https://www.sarvam.ai/blogs/sarvam-m>
- [24] Sarvam AI. Introducing Indus. 20 February 2026. <https://www.sarvam.ai/blogs/introducing-indus>
- [25] Sarvam AI. Sarvam Vision. 5 February 2026. <https://www.sarvam.ai/blogs/sarvam-vision>
- [26] Sarvam AI. Sovereign AI partnerships with Indian states. 8 February 2026. <https://www.sarvam.ai/blogs/partnerships-with-indian-states>
- [27] Sarvam AI. Series B announcement. 15 June 2026. <https://www.sarvam.ai/announcing-series-b>
- [28] Tech Mahindra. Project Indus launch. 28 June 2024. <https://www.techmahindra.com/insights/press-releases/tech-mahindra-launches-project-indus-large-language-model-llm/>
- [29] Tech Mahindra. The Indus Project. <https://www.techmahindra.com/makers-lab/indus-project/>
- [30] Tech Mahindra. Integrated Annual Report 2023 to 2024. <https://insights.techmahindra.com/investors/tml-integrated-annual-report-fy-2023-2024.pdf>
- [31] Tech Mahindra. Project Indus white paper. <https://www.techmahindra.com/insights/whitepapers/indus-project-indias-own-large-language-model/>
- [32] AIKosh. Param2 17B model card. https://aikosh.indiaai.gov.in/home/models/details/param2_17b_a2_4b_thinking.html
- [33] BharatGen. BharatGen programme. <https://bharatgen.com/>
- [34] Department of Science and Technology Government of India. National Mission on Interdisciplinary Cyber Physical Systems. <https://dst.gov.in/national-mission-interdisciplinary-cyber-physical-systems>
- [35] Ministry of Electronics and Information Technology. Digital Personal Data Protection Act 2023. <https://www.meity.gov.in/static/uploads/2024/06/2bf1f0e9f04e6fb4f8fef35e82c42aa5.pdf>
- [36] Ministry of Electronics and Information Technology. Digital Personal Data Protection Rules 2025. <https://www.meity.gov.in/documents/act-and-policies/digital-personal-data-protection-rules-2025-gDOxUjMtQWa?pageTitle=Digital-Personal-Data-Protection-Rules-2025>
- [37] Ministry of Electronics and Information Technology. Explanatory note to Digital Personal Data Protection Rules 2025. <https://www.meity.gov.in/writereaddata/files/Explanatory-Note-DPDP-Rules-2025.pdf>
- [38] Ministry of Electronics and Information Technology. Annual Report 2025 to 2026. <https://www.meity.gov.in/static/uploads/2026/04/46face7d48c8f6a97030f713ad5fdab4.pdf>
- [39] Reserve Bank of India. FREE-AI Committee Report. August 2025. https://m.rbi.org.in/scripts/bs_viewpublicationreport.aspx
- [40] Reserve Bank of India. Annual Report 2024 to 2025. <https://www.rbi.org.in/scripts/AnnualReportPublications.aspx?Id=1436>
- [41] Reserve Bank of India. Statement on responsible and ethical enablement of AI. 6 December 2024. https://www.rbi.org.in/Scripts/BS_ViewBulletin.aspx?Id=23057
- [42] National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework. <https://www.nist.gov/itl/ai-risk-management-framework>

- [43] National Institute of Standards and Technology. Generative AI Profile NIST AI 600-1.
<https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf>
- [44] International Organization for Standardization. ISO IEC 42001 artificial intelligence management systems.
<https://www.iso.org/standard/81230.html>
- [45] International Organization for Standardization. ISO IEC 23894 artificial intelligence risk management.
<https://www.iso.org/standard/77304.html>
- [46] International Organization for Standardization. ISO IEC 5259 data quality for analytics and machine learning. <https://www.iso.org/standard/81088.html>
- [47] OECD. OECD AI Principles. <https://oecd.ai/en/ai-principles>
- [48] UNESCO. Recommendation on the Ethics of Artificial Intelligence.
<https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence>
- [49] World Intellectual Property Organization. Artificial intelligence and intellectual property.
<https://www.wipo.int/en/web/frontier-technologies/artificial-intelligence/index>
- [50] MLCommons. MLPerf Inference benchmarks. <https://mlcommons.org/benchmarks/inference-datacenter/>

About the Author

Chennakeshav (CK) is a corporate finance and investment banking executive with 25+ years of global experience in deal origination, structuring and execution across M&A, growth capital and corporate strategy. He has led value-creation mandates for founders, corporates and funds — bridging the boardroom view to hands-on execution and close.

His career spans Morgan Stanley, HSBC, Lloyds Banking Group, EWEC, ADQ portfolio companies and Emirates Growth Fund, across TMT, real estate, fintech, deeptech, cleantech, infrastructure and energy. He has partnered with C-suite leaders, private equity and venture funds, sovereign wealth funds and family offices to finance complex fund raises and scale-up ventures, and has led M&A due diligence, post-merger integration and business-transformation initiatives to create value.

At Matchpoint Partners he is Managing Partner, leading the firm's corporate finance, M&A and capital-raising practice. He holds an MBA from London Business School, an engineering degree from VTU and a Master of Laws (LLM, in progress) from UCL London.

An active start-up mentor, CK mentors at Techstars, DIFC FinTech Hive, Startup Grind, Founder Institute and IN5, serves as Entrepreneur Mentor in Residence (EMiR) at London Business School, and judges the Entrepreneurship World Cup.

<https://www.linkedin.com/in/ckadya/>

<https://www.matchpoint-partners.com/team/ck-adya.html>

This paper is part of a continuing series on the structure of private and alternative markets. The views expressed are the author's own. The paper is for information only, describes market structure in general terms, and does not constitute investment, legal, tax or regulatory advice or a recommendation in respect of any security, vehicle or counterparty.