Spotting the Fake, Proving the Real: Valuing Deepfake-Defence Firms

An Acquisition and Valuation Framework for Standards, Customer Workflows, Adversarial Decay and Evidence Durability

Authored by Chennakeshav Adya

Independent Researcher

September 2026

Abstract

Synthetic image, audio, video and text systems can support legitimate creative and commercial uses while also lowering the cost of impersonation, fabricated evidence and manipulated communications. Buyers evaluating deepfake-defence firms therefore encounter a market with urgent customer needs, rapidly changing generation methods and multiple technical approaches. A detector can perform strongly on a familiar benchmark and weaken against a new generator, compression method or distribution channel. A provenance system can confirm that signed metadata remains intact without determining whether the recorded assertion is truthful. A watermark can help identify supported content yet disappear after transformation or be absent from content created outside the participating ecosystem.[1][2][3][4]

This paper develops an acquisition and valuation framework for deepfake-defence firms. The proposed unit of value is a verified authenticity decision delivered inside a customer workflow at complete cost. The framework tests six connected capabilities: provenance, watermark and label handling, forensic detection, identity and liveness verification, case orchestration, and accountable human review. It links performance evidence to customer cohorts, contract scope, response time, analyst effort, regulatory obligations, gross margin, renewal and cash.

NIST's 2024 synthetic-content report describes provenance, labelling, watermarking and detection as complementary transparency approaches. NIST's OpenMFC programme measures manipulation and deepfake detection under controlled evaluation, while the 2026 Guardians of Forensic Evidence programme emphasises generalisation, dirty post-processed evidence, continuous reassessment and operational validation. DARPA's Semantic Forensics programme adds detection, attribution and characterisation. C2PA Specifications 2.4 provide a current technical framework for cryptographically bound content provenance. The European Union's Article 50 transparency obligations became applicable on 2 August 2026, supported by a final Code of Practice covering marking, detection and labelling.[1][2][3][4][5][6]

A hypothetical acquisition illustrates a firm serving financial institutions, media platforms, public-sector bodies and enterprise communications teams. Every revenue, customer, performance, cost, probability and valuation figure in the example is a management assumption created only to demonstrate the method. It is neither a forecast nor a market benchmark.

The analysis concludes that a buyer should price evidence durability, workflow adoption and response economics before assigning value to headline accuracy. Six figures and seven tables convert the framework into diligence tests, a valuation bridge, consideration protection and a 180-day integration programme.

Cybersecurity, privacy, AI governance, communications, evidence, employment, competition, foreign investment, accounting, tax, insurance and securities decisions require current advice from qualified

specialists in each relevant jurisdiction. This paper provides general information for professional audiences and does not provide legal, regulatory, technical, accounting, tax or investment advice.

JEL Classification: G24, G34, L86, O32, O33

Keywords: deepfake defence, synthetic media, content provenance, media forensics, artificial intelligence, cybersecurity M&A, valuation, integration

1. Define the acquisition decision

The board should identify the decision that customers pay the target to improve. Deepfake-defence firms can sell media-forensics software, provenance infrastructure, watermark detection, identity verification, liveness checks, fraud orchestration, managed investigation, brand protection or expert evidence services. Each model has a different evidence chain, customer buyer, cost structure and liability perimeter. A transaction thesis that groups all of them as synthetic-media detection can misstate both strategic fit and recurring economics.

The acquisition memorandum should specify the intended source of value. A buyer may seek proprietary detection research, access to regulated customers, a provenance network, workflow integration, scarce forensic specialists, channel distribution or a platform for consolidation. Each source needs an observable test. Technical performance should be reproduced on representative data. Customer value should appear in adoption, decisions, response outcomes, renewal, pricing and cash. Product options should remain separately identified until rights, performance and demand are evidenced.

The board should compare acquisition with licensing, commercial partnership, minority investment and internal development. Ownership can matter where value depends on coordinated control of models, evaluation data, customer integrations, forensic personnel and trust infrastructure. A narrower arrangement can carry less risk where interoperability or channel access creates the main benefit.

Evidence timing should shape price and terms. Some tests can be completed before signing using controlled data. Others require customer permission, production workflows or post-close access. The buyer can use conditions, retained consideration, warranties, covenants and earn-outs to address gaps. The transaction team should state which evidence supports the base price, which evidence supports contingent value and which capability remains an option.

Figure 1. Authenticity-to-value evidence chain

The proposed chain connects media evidence to an accountable customer decision, operational response and collected cash.

2. Define the unit of value

The proposed unit of value is a verified authenticity decision delivered inside a customer workflow at complete cost. A decision may confirm that content carries valid provenance, indicate that evidence is

likely manipulated, escalate uncertainty to a human examiner, block a payment instruction, protect an account, label content, preserve evidence or trigger further verification. The decision should have a recorded purpose, population, threshold, evidence basis, owner and permitted action.

Complete cost includes ingestion, storage, model inference, third-party licences, cryptographic trust services, watermark services, evaluation data, model retraining, analyst labour, quality review, customer integration, support, legal review, incident handling, insurance, compliance and working capital. A product can appear highly scalable while expert investigators resolve difficult cases outside the reported service cost. The acquisition model should capture every activity required to reach the promised outcome.

Accuracy alone is an incomplete unit. Performance varies by modality, generator, editing chain, language, device, compression, duration, demographic characteristic, channel and operating threshold. A customer also needs latency, evidence presentation, auditability and a route for uncertain cases. A model with a strong area-under-curve score can still produce unacceptable false positives at the low false-alarm rate required for payment approval or public attribution.

The buyer should therefore measure verified decisions per complete case cost, with separate controls for error consequence. This denominator aligns technical, operational and financial diligence. It also provides a basis for comparing software, managed service and expert-evidence models without assuming that every alert or API call creates equivalent value.

3. Map the capability stack

Deepfake defence is a layered control problem. Provenance records can show the source and editing history asserted in a signed manifest. Watermarks and labels can signal synthetic origin for participating systems. Forensic models can analyse content where provenance is absent, damaged or untrusted. Identity and liveness controls can test whether a person is present and authorised. Workflow systems can combine signals, apply policy and preserve evidence. Human examiners can handle ambiguity, consequence and legal context.[1][5][7]

These layers should not be collapsed into a single accuracy claim. C2PA explains that Content Credentials provide tamper-evident provenance and do not make a value judgement about whether provenance assertions are true. Detection systems face distribution shift and adversarial countermeasures. Identity controls address a narrower question than media authenticity. Human review adds judgement and accountability while increasing cost and capacity risk.[5][8]

The diligence team should map each product module to its claim, evidence, failure mode and customer action. The map should identify proprietary, licensed, open-source and customer-supplied components. It should include certificate and trust-list dependencies, model providers, reference databases, biometric vendors, case-management systems and cloud services. Replacement rights and fallback paths should be recorded.

Table 1. Deepfake-defence capability stack

Layer	Customer question	Core evidence	Principal limitation
Provenance	What source and edit history are asserted?	signed manifest, certificate, ingredient chain	assertion quality and ecosystem coverage
Watermark and label	Was participating content marked?	detectable signal and issuer record	transformation, removal and non-participation
Forensic detection	Is content likely generated or manipulated?	calibrated score and supporting features	generalisation and adversarial decay
Identity and liveness	Is the claimed person present and authorised?	challenge, biometric and device evidence	scope, bias, privacy and presentation attacks

Layer	Customer question	Core evidence	Principal limitation
Workflow	What action follows the evidence?	policy, case history and audit trail	integration and inconsistent customer practice
Human review	How is uncertainty resolved?	examiner method, peer review and report	cost, capacity and judgement variation

Each layer answers a different question and creates a different diligence requirement.

4. Build the authenticity-to-outcome ledger

The ledger should connect each sampled item to its source, media type, acquisition channel, compression history, provenance state, watermark state, model versions, forensic scores, identity evidence, analyst review, decision, customer action, dispute and financial record. The purpose is traceability across product, operations and value.

Negative and uncertain cases belong in the ledger. False accusations of manipulation can damage customers and subjects. Missed manipulations can enable fraud or misinformation. Inconclusive results can be the correct outcome when evidence is weak. A data room containing only successful demonstrations cannot support a population-level conclusion.

The buyer should sample across customer cohorts and consequences. A bank validating a payment instruction, a news organisation checking incoming footage, a platform applying a label and a court-facing examiner have different thresholds and evidence requirements. The ledger should preserve these differences rather than averaging them into one score.

Finance should link the same cohorts to contracted revenue, usage, service effort, credits, claims, renewal, expansion and collections. This reveals whether performance converts to customer value and whether difficult cases consume unpriced labour. It also exposes concentration in a generator, platform, channel or regulatory requirement that can change quickly.

5. Design a representative benchmark

NIST OpenMFC separates manipulation detection, deepfake detection and localisation tasks and uses measures including receiver-operating-characteristic curves, area under the curve and correct detection rate at a specified false-alarm rate. The 2026 Guardians programme extends the operational emphasis to newer generators, post-processed evidence, representative casework and continuous reassessment.[2][3]

The buyer should construct a benchmark that reflects the target's contracted use. It should include authentic, edited, generated and manipulated content; known and previously unseen generators; multiple languages and accents; different devices and capture conditions; common social-media and messaging transformations; and deliberate anti-forensic changes. Training, development and final challenge sets should be separated.

Performance should be reported at operating thresholds. A customer that processes millions of benign items may require a very low false-alarm rate. A forensic triage service can accept a different threshold when human review follows. The buyer should measure calibration, abstention and confidence stability as well as ranking performance.

Benchmark governance matters. The target should record dataset rights, provenance, demographic composition, contamination checks, label process, model access and evaluation history. Repeated tuning on a visible test set can create apparent performance that does not generalise. Independent challenge administration and sealed final sets can reduce this risk.

The benchmark should also distinguish component tests from end-to-end tests. A detector can classify a clip correctly while the surrounding workflow loses the file, applies the wrong threshold or fails to present

the reason to an approver. End-to-end tests should begin at the customer entry point and continue through ingestion, normalisation, scoring, provenance validation, review, decision, response and audit export. The buyer should record failures caused by integration, data quality and operations alongside model errors.

Sample design should follow the commercial exposure. If most revenue comes from short mobile audio submitted through messaging applications, a large benchmark of high-resolution studio video provides limited valuation evidence. The team should weight results by the target's current and contracted customer population, then present unweighted results as a technical cross-check. New-market scenarios should use separate assumptions because the future population can differ materially from the installed base.

Table 2. Representative acquisition benchmark

Dimension	Evidence population	Measure	Acquisition warning
Modality	image, audio, video and mixed media	calibrated performance by modality	one strong modality presented as a platform result
Generator shift	known and unseen generators	time-separated challenge performance	training and test contamination
Transformation	compression, crop, blur, re-recording and transcription	performance after realistic processing	pristine laboratory files only
Identity	face, voice and liveness scenarios	error rates by cohort and attack	aggregate rate conceals subgroup weakness
Operations	channel-specific case workflow	accepted decision, latency and rework	API score without customer action
Evidence	reports and audit trail	completeness and reproducibility	score cannot be explained or reproduced

The benchmark separates capability, operating threshold, robustness and workflow outcome.

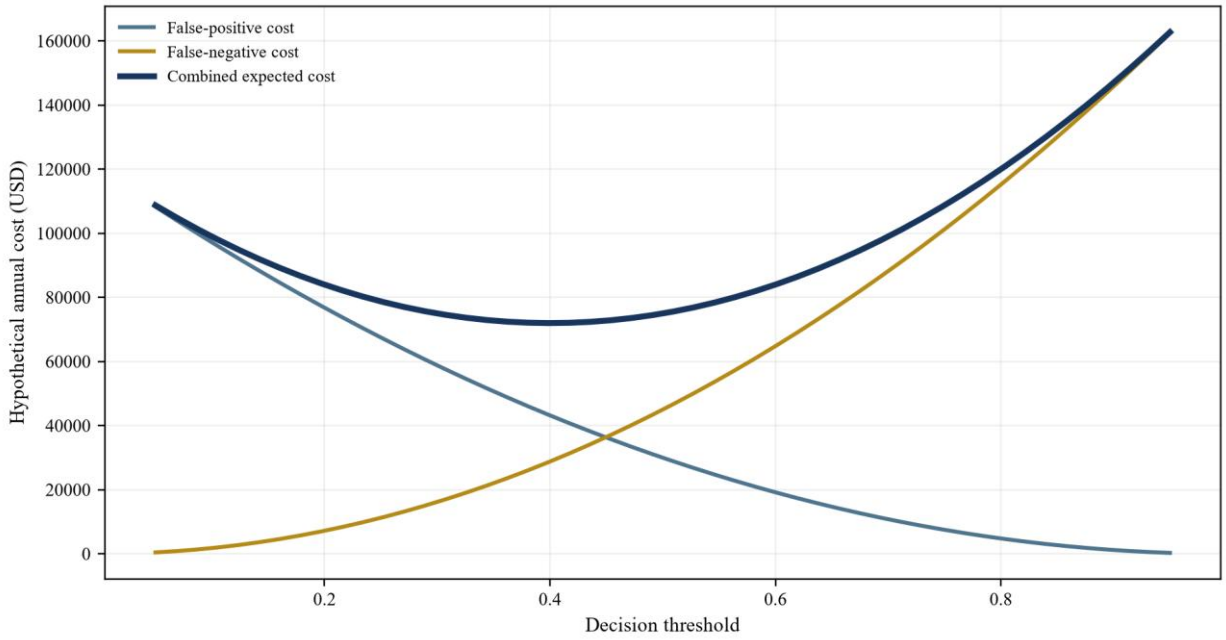
6. Measure false positives and false negatives

Error cost should drive the operating point. A false positive can block a legitimate payment, mislabel authentic journalism, reject a customer or undermine evidence. A false negative can permit impersonation, fraud or manipulated content. The buyer should identify who bears each consequence and how contracts allocate responsibility.

Detection reports should include confusion matrices at relevant thresholds, receiver-operating-characteristic and precision-recall curves, calibration, abstention and cohort results. Prevalence matters. A test set balanced between real and fake content can overstate positive predictive value in a production stream where manipulated content is rare. The buyer should recompute expected workload using observed customer prevalence.

Human review can reduce some errors while creating queue and consistency risk. The diligence team should compare automated, assisted and examiner-only decisions on controlled cases. Disagreements should be resolved through documented adjudication. The analysis should record whether the system changes the final decision, time, evidence quality or reviewer confidence.

Figure 2. Hypothetical error-cost curve



The illustration shows how customer consequence can shift the preferred operating threshold. Values are management assumptions for method demonstration.

7. Test adversarial decay

Adversarial decay is the reduction in useful performance as generation methods, editing tools, evasion techniques and distribution channels change. NIST's Guardians programme explicitly targets generalisation to newer methods and dirty post-processed evidence. NIST's adversarial-machine-learning taxonomy describes evasion, poisoning, privacy and misuse attacks across predictive and generative systems.[3][9]

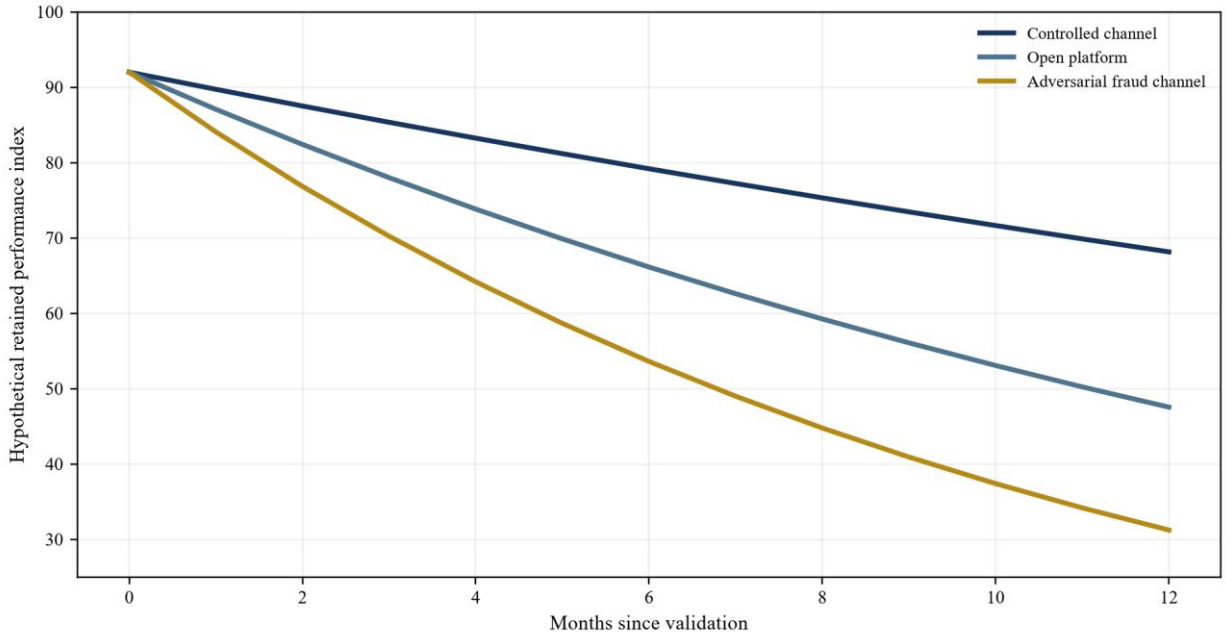
The buyer should build time-separated challenge sets. Models can be frozen at defined dates and tested against content produced later. The team should measure performance by generator family, transformation and channel, then record the research and engineering effort required to restore the approved operating point. The result is an evidence half-life rather than a permanent accuracy claim.

Decay also occurs in provenance and watermark systems. Trust lists, certificates, manifests, soft bindings, issuer participation and verification software require continuing maintenance. Compatibility changes can improve security while creating migration obligations. The acquisition model should include the people, infrastructure and governance required to keep the evidence chain current.

Research throughput should be measured as an operating process. The target should show how it discovers new generation methods, acquires lawful evaluation material, labels evidence, prioritises threats, changes models, validates releases and informs customers. The buyer should examine cycle time from a material new threat to an approved customer release, together with regression failures, emergency patches and analyst workarounds. A fast release process creates limited value when it repeatedly damages established performance.

The diligence team should separate durable platform capability from model-specific advantage. Data pipelines, experiment controls, evaluation harnesses, customer feedback, model monitoring and expert teams can support repeated adaptation. A narrow classifier trained on a transient artefact can lose value quickly. Valuation should recognise the process that renews evidence as well as the current model result, provided that the process is documented, staffed and reproduced during diligence.

Figure 3. Hypothetical evidence-durability curve



The curves illustrate different rates of performance decay between validation cycles. Values are management assumptions for method demonstration.

8. Evaluate provenance and Content Credentials

C2PA Specifications 2.4 describe cryptographically verifiable content provenance, including manifests, assertions, ingredients, signatures, trust and soft binding. Provenance can show who or what signed an assertion and whether the asset remains associated with it. The buyer should test creation, embedding, sidecar storage, verification, trust-list handling, revocation, ingredient chains and transformed-asset recovery.[5][10]

Coverage is a commercial question. Provenance is most valuable when capture devices, generation tools, editors, publishers, platforms and customer workflows participate. The buyer should measure the share of relevant content with usable credentials, the share that survives each channel and the actions customers take when credentials are present, absent or invalid.

Trust governance should receive transaction diligence equal to model governance. Certificate issuance, key custody, identity proofing, revocation, timestamping and software update practices can determine whether a credential is useful. Customer contracts should state what the system verifies and what remains outside scope.

The target's intellectual property should be separated from ecosystem participation. Implementation expertise, workflow integration, trust services and proprietary recovery methods may create value. Access to an open standard alone does not establish a moat. The buyer should test switching cost, interoperability and customer ownership of manifests and records.

Interoperability tests should use credentials created by multiple conforming tools and validated in the target's supported environments. The team should test malformed manifests, revoked credentials, expired certificates, missing ingredients, conflicting assertions and content that has been transformed or stripped of metadata. The product should present validation status and limitations consistently. Customer policy should distinguish a valid signature from trust in the signer and from confidence in the underlying event.

9. Evaluate watermarking and labelling

Watermarking can help identify supported synthetic content through signals embedded at generation or editing. Labels can communicate artificial origin to users. NIST identifies watermarking and labelling among the technical approaches to synthetic-content transparency, while the EU's final 2026 Code of Practice addresses marking, detection and deepfake labelling under Article 50.[1][6]

The diligence programme should test detectability after resizing, cropping, compression, transcoding, re-recording, mixing, translation and common editing. It should measure false attribution and the ability to distinguish multiple issuers or model families where the product makes those claims. Watermark security, key management and collusion risk should be included.

Commercial value depends on workflow adoption. A watermark detector used only during demonstrations has little recurring value. The buyer should trace the signal into moderation, fraud review, publishing, communications and evidence-preservation systems. It should measure response time, human review and customer policy after detection.

Non-participating content remains a central limitation. Watermark absence does not prove authenticity, and presence should be interpreted within the issuer and threat model. Product language, training and contracts should prevent customers from converting a partial signal into an unsupported conclusion.

10. Test multimodal and cross-channel performance

Deepfake campaigns can combine synthetic voice, manipulated video, fabricated documents, compromised accounts and social engineering. A media-only detector can miss the contextual evidence that makes the event actionable. The buyer should test how the platform links content, identity, device, account, communication and transaction signals.

Audio deserves separate treatment because codecs, background noise, short clips, multilingual speech, voice conversion and replay affect performance. Video introduces temporal consistency, lip synchronisation, frame selection and compression. Images create different provenance and localisation questions. Text and document workflows add authorship, metadata and semantic consistency problems.

Cross-channel orchestration can create a defensible workflow advantage when it reduces customer effort and improves decisions. It can also increase integration cost and data-rights exposure. The buyer should identify which correlations are production features, which depend on customer data and which remain research prototypes.

11. Assess customer workflows

The product should be evaluated where the decision occurs. Financial institutions can use authenticity checks in payment approval, account recovery, remote onboarding and executive impersonation response. Media organisations can use them in source verification and publishing. Enterprises can protect executive communications and brand channels. Platforms can support labelling, moderation and appeals. Public agencies and forensic laboratories can use validated tools within evidence procedures.[3][11][12]

For each workflow, the buyer should document the trigger, evidence available, service level, decision owner, permitted actions, escalation, audit trail and redress. The product's contribution should be isolated from other controls. A reduction in fraud losses can reflect stronger callbacks, transaction limits or account controls as well as detection.

Adoption should be measured beyond licences. Useful measures include active protected workflows, eligible events analysed, cases reviewed, decisions changed, actions completed, appeals, service credits and verified loss avoidance. Customer references should be reconciled to system and contract records.

The buyer should examine the counterfactual. A customer may have bought the product after a public incident while continuing to rely on manual callbacks and existing identity controls. Interviews should establish which decisions changed, what alternative process would have occurred, and which measured benefit can reasonably be attributed to the target. Loss avoidance requires special care because prevented events are difficult to observe. The analysis should use documented attempted incidents, controlled simulations and customer-approved attribution methods.

Workflow depth can create retention when the product becomes embedded in policy, evidence handling and case systems. It can also create implementation liabilities. The diligence team should measure time to deployment, customer configuration, training, change management, integration maintenance and the share of customers operating below contracted scope. Services required to make the software useful should remain visible in margin and capacity models.

Table 3. Customer workflow and outcome map

Customer workflow	Decision	Evidence required	Outcome measure
Payment approval	release, hold or verify instruction	identity, channel, content and callback evidence	prevented loss, delay and false hold
Remote onboarding	accept, reject or escalate applicant	liveness, identity, document and device evidence	fraud, completion and review cost
News verification	publish, label, investigate or reject	provenance, forensic and source evidence	correction, speed and confidence
Platform moderation	label, limit, remove or appeal	content, policy and provenance evidence	exposure, appeal and error rates
Enterprise communications	trust, challenge or isolate message	account, voice, video and context evidence	incident containment and disruption
Forensic examination	report finding and limitation	validated method, chain of custody and peer review	reproducibility and case acceptance

Value arises when technical evidence changes a controlled customer decision.

12. Evaluate model governance and explainability

The buyer should obtain a complete model inventory covering modality, purpose, architecture, training data, external components, versions, deployment, thresholds, known limitations and owners. Changes should be linked to evaluation evidence and customer release records. NIST AI RMF and the Generative AI Profile provide useful structures for governing, mapping, measuring and managing AI risk.[13][14]

Explainability should be tied to the user and decision. An investigator may need localisation, signal contribution, provenance validation and alternative explanations. A customer approver may need a concise evidence summary and limitation statement. A court-facing expert requires method validation, chain of custody and reproducibility.

The system should support abstention and uncertainty. Forced binary answers can create customer harm when evidence is weak or outside the validated population. The buyer should review how uncertainty affects workflow, service levels and reported performance.

Model-change governance should connect research evidence to commercial release. The target should define material change, required regression scope, approval authority, customer notice and rollback conditions. A threshold change can alter workload and customer outcomes even when model weights remain unchanged. External model or API updates can create similar effects. The buyer should sample releases and trace each one from development evidence through approval, deployment, monitoring and customer communication.

Documentation should remain usable by product teams, investigators, sales teams and customers. Technical limitations hidden in research notes can disappear from proposals and operating procedures. The diligence team should compare model cards, validation reports, product documentation, training materials, statements of work and customer-facing claims. Differences should feed the liability assessment and remediation plan.

13. Protect data, rights and chain of custody

Training, evaluation and production data can include faces, voices, communications, identity documents, confidential media and criminal evidence. The buyer should map legal basis, consent, licence, purpose, geography, retention, security, deletion, onward transfer and customer restrictions. Rights should cover the intended post-close use, including model improvement and cross-customer learning where applicable.

Chain of custody matters when outputs support investigations, disputes or proceedings. The platform should preserve acquisition method, original hashes, transformations, model and software versions, analyst actions, timestamps and exports. Records should remain reproducible after product updates.

Table 4. Data, model and evidence-rights register

Asset	Rights and control evidence	Operational test	Valuation treatment
Training data	licence, consent, provenance and restrictions	reproduce authorised dataset	value only transferable permitted use
Evaluation data	independent labels and contamination controls	rerun sealed benchmark	support evidence durability
Production media	contract, purpose and retention	trace sampled case	recognise only authorised workflow value
Models	ownership, third-party terms and dependencies	rebuild or restore version	separate owned and licensed capability
Credentials and keys	issuer, custody, trust and revocation	validate and revoke test credential	price trust service and liability
Case evidence	chain of custody and audit rights	reproduce report	support regulated and expert workflows

The register connects legal rights, technical lineage and valuation use.

14. Test security and abuse resistance

The target is itself an adversarial system. Attackers can probe thresholds, submit crafted media, poison feedback, steal models, compromise keys, flood review queues or manipulate analyst context. NIST's adversarial-machine-learning taxonomy and C2PA security guidance provide useful starting points for threat modelling.[9][15]

The buyer should commission testing across model, API, application, identity, certificate, data and operational layers. Rate limits, anomaly detection, access control, secure development, incident response, secrets management and recovery should be reviewed. Customer-supplied evidence should be treated as potentially hostile.

Abuse controls also govern customers. A forensic tool can be used for surveillance, unsupported attribution or collection of biometric data. The target should define acceptable use, high-risk customers, escalation and termination. Revenue that depends on prohibited or weakly controlled use should be excluded from the base case.

15. Assess regulatory and evidence exposure

The regulatory perimeter varies by product and jurisdiction. The EU AI Act's Article 50 transparency obligations apply to marking and labelling of certain AI-generated content and deepfakes from 2 August 2026. The final Code of Practice provides a voluntary compliance route for signatories while legal obligations remain in the Act. Data protection, biometric, platform, communications, consumer-protection and evidence rules can also apply.[6][16][17]

US authorities have addressed impersonation and AI-generated voice harms through the FTC's government and business impersonation rule and the FCC's treatment of AI-generated voices under the Telephone Consumer Protection Act. FBI and IC3 alerts document criminal use of AI-generated voice and altered media in impersonation and fraud campaigns.[11][12][18]

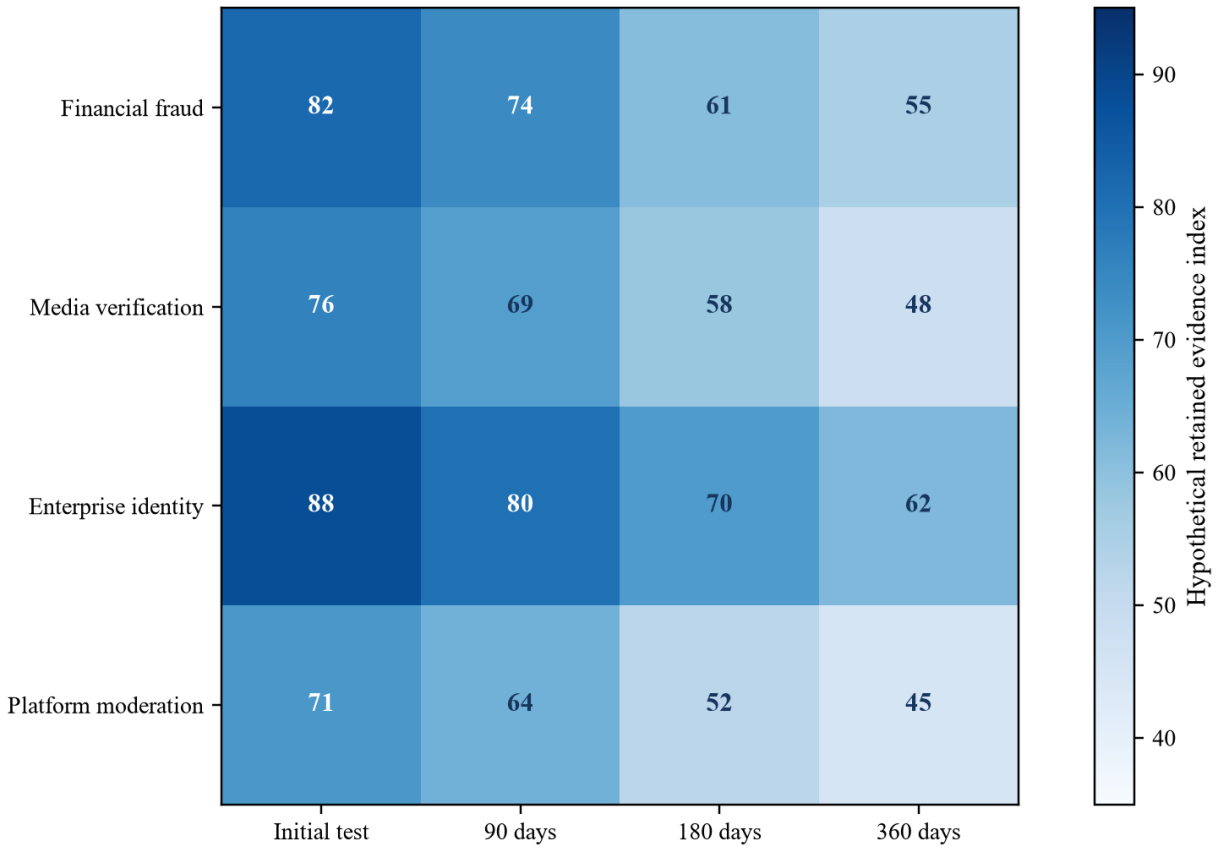
The buyer should obtain a jurisdiction-by-product matrix and current specialist advice. Product claims, customer actions, retention, reporting and human review should match the relevant obligations. Regulatory demand can create revenue while also increasing validation, audit and liability cost.

16. Analyse customer cohorts and contracts

Customers should be grouped by workflow, consequence, modality, integration depth, geography, channel and service model. Revenue concentration by platform, generator or regulatory requirement should be visible. Cohorts should show contracted recurring revenue, usage, deployment, review effort, gross margin, renewal, expansion, claims and cash.

Contract review should cover performance commitments, permitted use, data rights, audit, security, incident response, indemnity, liability caps, insurance, service credits, evidence retention, model changes and termination. Marketing claims should be reconciled with enforceable scope.

Figure 4. Customer cohort evidence matrix



The matrix illustrates how workflow adoption and evidence durability can be reviewed together.

17. Rebuild complete delivery economics

Gross margin should be rebuilt from customer and case records. Costs include cloud processing, model inference, storage, trust infrastructure, third-party data, certificates, analyst review, research, evaluation, customer integration, support, incident response and insurance. Shared research cost should be allocated explicitly.

Usage growth can lower unit cost through scale while increasing review and infrastructure cost. High-consequence customers may require dedicated environments, expert availability and stronger evidence retention. The buyer should separate software margin from managed-service and expert-evidence margin.

Cash conversion requires review of billing triggers, minimum commitments, overages, professional services, credits, disputes and collection. A large contracted value can coexist with slow deployment and limited usage. Deferred revenue and customer-funded implementation should be reconciled to remaining obligations.

18. Build a hypothetical acquisition case

Assume a target has 145 customers, annual recurring revenue of USD 42 million and trailing revenue of USD 47 million. Sixty percent of recurring revenue comes from financial-fraud and identity workflows, twenty-five percent from media and platform workflows, and fifteen percent from enterprise and public-sector investigation. Reported gross margin is 76 percent.

Management estimates that complete case-review, evaluation-data and trust-infrastructure costs reduce normalised gross margin to 68 percent. Twelve customers representing USD 9 million of recurring revenue depend on one distribution partner. A further USD 6 million depends on a model feature whose performance on the buyer's sealed unseen-generator set falls below the approved operating point.

The buyer reproduces acceptable performance for USD 27 million of recurring revenue, confirms contract and data rights for USD 25 million, and validates current workflow adoption for USD 23 million. These figures are management assumptions for method demonstration. They should not be used as a forecast, market multiple or statement about a real company.

Table 5. Hypothetical acquisition evidence bridge

Item	Hypothetical amount	Evidence status	Treatment
Reported recurring revenue	USD 42m	contract list	starting point
Reproduced-performance cohort	USD 27m	sealed benchmark and case sample	eligible for evidence-backed base
Transferable-rights cohort	USD 25m	contracts and rights register	eligible after legal confirmation
Active-workflow cohort	USD 23m	production and customer records	highest-confidence recurring base
Partner-concentrated revenue	USD 9m	distribution agreement	concentration stress and retention
Feature-dependent revenue	USD 6m	failed unseen-generator threshold	remediation or contingent value
Normalised gross margin	68%	complete cost rebuild	use in cash-flow case

Every figure is a management assumption created solely to demonstrate the valuation method.

19. Stress the acquisition model

The model should test performance decay, customer false-positive tolerance, partner loss, model-provider price changes, certificate or trust-list changes, regulatory remediation, expert attrition, major abuse incidents and delayed integration. Each scenario should flow through usage, review effort, credits, renewal, revenue, margin, investment and cash.

Stress tests should distinguish temporary remediation from structural impairment. A detector that can recover through routine retraining differs from one that requires new data rights or architecture. A provenance service with diversified issuers differs from one dependent on a single platform. Customer workflows with layered verification differ from those relying on a single score.

The board should receive downside cases that preserve operating reality. Cutting research or evaluation spend can improve near-term margin while accelerating evidence decay. Eliminating reviewers can raise throughput while increasing errors. Integration savings should be released only after controlled cohorts meet performance and service thresholds.

Sensitivity should be expressed through operating drivers that management can monitor after closing. These include the share of content within the validated population, refresh cycle time, reviewer minutes per case, partner concentration, certificate-service cost, customer activation and renewal after an adverse event. The investment committee can then link each assumption to an owner, reporting cadence and intervention threshold.

This creates a measurable post-close control system.

20. Value the firm in layers

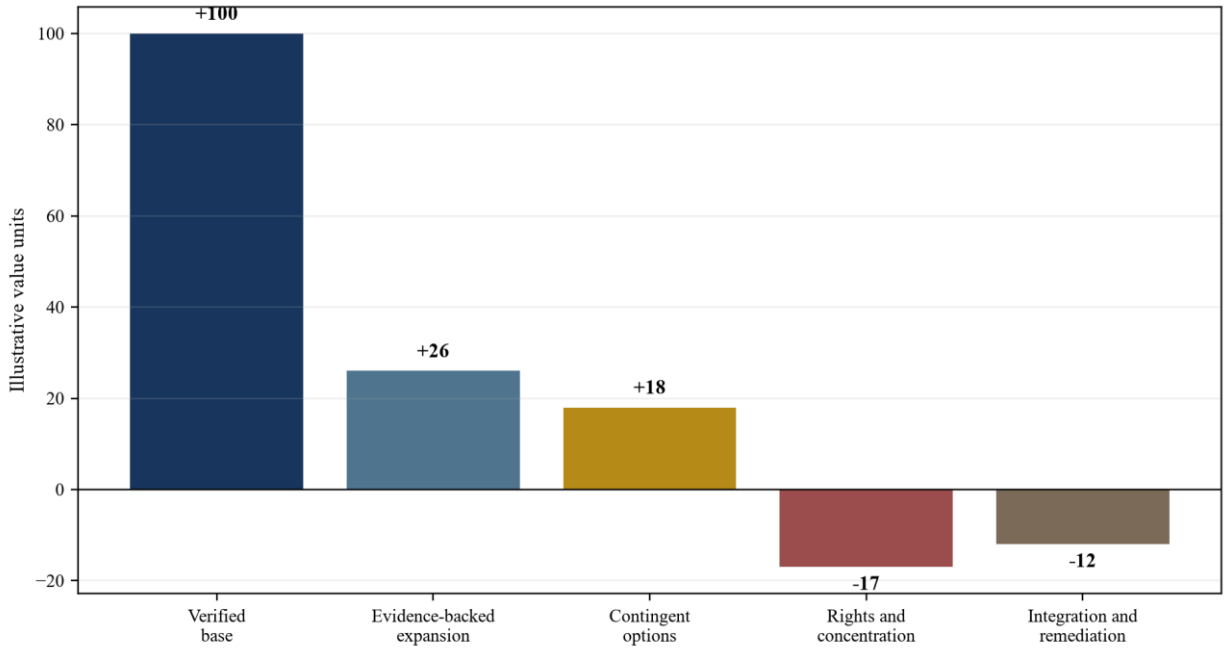
The base layer is cash flow from active customer workflows with reproduced performance, transferable rights, normalised cost and acceptable concentration. A second layer can recognise evidence-backed expansion where deployment prerequisites are complete. A third layer can recognise contingent product or ecosystem options through probability-weighted scenarios or milestone consideration.

Revenue multiples can provide a market cross-check after business models and evidence quality are normalised. A managed forensic service with expert-heavy delivery, a high-volume identity API and a provenance network can report recurring revenue while carrying materially different margins, reinvestment, concentration and liability. The comparable set should be explained rather than treated as a direct pricing rule. Cash-flow scenarios should reflect research renewal, evaluation, integration and trust-service costs.

The cost approach can inform the value of software, datasets and technical reconstruction, yet historical spending does not establish usefulness or legal rights. The income approach can isolate customer relationships or technology contributions when assumptions are supported. Relief-from-royalty, excess-earnings and replacement-cost methods require careful separation of contributory assets and double counting. The board should reconcile commercial price, financial reporting allocation and downside protection without treating them as the same exercise.

Intangible assets may include software, models, datasets, customer relationships, contracts, trademarks and in-process research. IFRS 3 and IAS 38 address identifiable intangible assets acquired in a business combination, while IFRS 13 and IVS provide valuation principles. The transaction model should keep accounting allocation separate from the commercial decision while reconciling both.[19][20][21][22]

Figure 5. Evidence-layered valuation bridge



The illustration separates verified value, evidence-backed value, contingent options and deductions.

Table 6. Valuation and consideration design

Value component	Required evidence	Valuation approach	Protection mechanism
Active recurring base	reproduced performance, rights, adoption and cash	normalised cash flow and market cross-check	closing adjustment and warranties
Expansion	contracted customers and completed deployment gates	probability-weighted cohort cash flow	staged payment
Research advantage	time-separated benchmark and refresh economics	excess return or cost-to-recreate cross-check	retention and milestone payment
Provenance network	issuer participation and workflow usage	network cohort economics	adoption milestone
Product option	prototype and documented route to market	scenario value	earn-out or retained equity
Liabilities	rights, claims, remediation and concentration	expected cost and downside scenario	indemnity, escrow or price reduction

Consideration should follow the strength and timing of evidence.

21. Design transaction terms and integration

Representations should cover ownership, training and evaluation rights, model inventory, provenance and certificate operations, security, privacy, customer contracts, performance claims, incidents and disputes. Disclosure schedules should identify third-party models, data, open-source components, trust dependencies and key personnel.

Earn-outs should use measures that management can influence and both parties can audit. Measures can combine retained recurring revenue, normalised gross margin, active workflow adoption, performance on defined time-separated challenge sets and customer outcomes. Dataset, threshold, model version, population, review authority and dispute process should be agreed before closing.

Integration should preserve evidence. Models, datasets, certificates, keys, logs, case histories and customer configurations should remain controlled. Customer migration should use dual-run cohorts.

Research and forensic teams need retention, authority and a clear release process. Cost actions should follow verified capacity and performance.

22. Execute a 180-day programme

Days 1 to 30 establish control. The team confirms product and contract perimeter, model and data inventories, key custody, trust dependencies, customer cohorts and the authenticity-to-outcome ledger. It preserves benchmark sets, source evidence, versions and case records.

Days 31 to 75 reproduce performance. The team runs sealed cross-modality, unseen-generator and post-processing tests at customer thresholds. It samples production cases, human review and chain of custody. Finance rebuilds complete delivery cost and cash conversion.

Days 76 to 120 test durability and concentration. The team runs time-separated challenges, threat-model exercises and partner, model-provider and regulatory scenarios. It measures refresh effort, customer adoption and error consequence.

Days 121 to 180 finalise consideration protection and cohort integration. The board receives verified value, evidence-backed value, options, liabilities, first-year investment and release gates. Migration proceeds only after dual-run evidence meets approved thresholds.

Figure 6. Evidence-gated 180-day programme

The programme moves from control and reproduction to durability testing, transaction protection and cohort release.

Table 7. Evidence-gated transaction and integration plan

Gate	Evidence	Decision enabled	Failure response
Perimeter	product, contract and responsibility map	confirm acquisition scope	exclude or reprice unsupported scope
Performance	sealed representative benchmark	accept core capability	remediate, retain value or stop
Durability	time-separated and post-processing tests	recognise research advantage	reduce value and increase refresh investment
Rights and trust	data, model, certificate and customer rights	recognise transferable value	restrict use or reduce price
Customer outcomes	workflow adoption, error and cash cohorts	accept recurring base	adjust revenue and margin
Integration readiness	dual-run plan, key control and people coverage	release cohort migration	extend parallel operation

Gate	Evidence	Decision enabled	Failure response
Realised value	independently reconciled performance and cash	release retained consideration	defer or reduce payment

Each gate requires reviewable evidence before price, migration or cost action proceeds.

23. Decision and conclusion

Deepfake-defence firms should be valued through the durability of evidence they create inside real customer workflows. Headline accuracy can support technical review, yet transaction value requires representative populations, relevant operating thresholds, transferable rights, complete cost, adoption and accountable outcomes.

The control stack is complementary. Provenance, watermarking, forensic detection, identity checks, workflow orchestration and human review answer different questions. A buyer should identify the role and limitation of each layer and avoid assigning a universal authenticity claim to a partial signal.

Adversarial decay belongs in the valuation model. Time-separated challenge sets, dirty post-processed evidence, regression testing and refresh economics show how long a capability remains useful and what it costs to maintain. Research capacity is valuable when it produces reproducible performance and customer outcomes.

The transaction should maintain five connected records: the authenticity-to-outcome ledger, model and dataset inventory, rights and trust register, customer cohort model and evidence-durability schedule. These records allow technical, operational, regulatory and financial diligence to use the same evidence.

Price and integration should follow proof. Active recurring workflows with reproduced performance and transferable rights support the base. Evidence-backed expansion can support staged value. New modalities, ecosystems and product options remain contingent until adoption and economics are demonstrated. This structure gives buyers and sellers a practical basis for price, protection and execution in a market where threats and defences both change rapidly.

Frequently asked questions

What is the most useful unit for valuing a deepfake-defence firm?

Use a verified authenticity decision delivered inside a customer workflow at complete cost. The measure should connect media evidence, models and provenance signals to an accountable decision, customer action and cash.

Why is headline detection accuracy insufficient for valuation?

Performance changes with modality, generator, editing, channel, prevalence and threshold. A buyer needs cohort-level false positives, false negatives, calibration, abstention, latency and workflow outcomes.

How should a buyer test adversarial decay?

Freeze model versions and evaluate them against time-separated content from newer generators, transformations and channels. Measure performance loss, recovery time and the complete cost of restoring the approved operating point.

How should Content Credentials be valued?

Value the target's implementation, trust services, recovery capability, customer adoption and workflow integration. Ecosystem coverage, issuer trust, key custody, interoperability and customer actions determine commercial value.

Does the absence of a watermark prove that content is authentic?

No. Content can come from a non-participating system, lose a signal through transformation or use another marking method. Watermark evidence should be interpreted within the issuer and threat model.

What should a deepfake benchmark include?

Include authentic, edited, generated and manipulated media across modalities, languages, devices, known and unseen generators, common post-processing, distribution channels and relevant operating thresholds.

What should an earn-out measure?

Measures can combine retained recurring revenue, normalised margin, active workflow adoption, performance on defined challenge sets and customer outcomes. Definitions and review authority should be agreed before closing.

When should customer migration proceed?

After representative dual-run cohorts demonstrate acceptable performance, evidence integrity, service levels, customer obligations and operational resilience under the buyer's target architecture.

References

[1] National Institute of Standards and Technology, Reducing Risks Posed by Synthetic Content, NIST AI 100-4. https://www.nist.gov/publications/reducing-risks-posed-synthetic-content-overview-technical-approaches-digital-content

[2] National Institute of Standards and Technology, Open Media Forensics Challenge. https://mfc.nist.gov/

[3] National Institute of Standards and Technology, Guardians of Forensic Evidence. https://www.nist.gov/programs-projects/guardians-forensic-evidence

[4] Defense Advanced Research Projects Agency, Semantic Forensics. https://www.darpa.mil/research/programs/semantic-forensics

[5] Coalition for Content Provenance and Authenticity, C2PA Specifications 2.4. https://spec.c2pa.org/specifications/specifications/2.4/index.html

[6] European Commission, Code of Practice on Transparency of AI-generated Content. https://digital-strategy.ec.europa.eu/en/policies/code-practice-ai-generated-content

[7] Coalition for Content Provenance and Authenticity, Guidance for Artificial Intelligence and Machine Learning. https://spec.c2pa.org/specifications/specifications/2.4/ai-ml/ai_ml.html

[8] Coalition for Content Provenance and Authenticity, C2PA Explainer. https://spec.c2pa.org/specifications/specifications/2.4/explainer/Explainer.html

[9] National Institute of Standards and Technology, Adversarial Machine Learning, NIST AI 100-2e2025. https://csrc.nist.gov/pubs/ai/100/2/e2025/final

[10] Coalition for Content Provenance and Authenticity, Content Credentials technical specification. https://spec.c2pa.org/specifications/specifications/2.4/specs/C2PA_Specification.html

[11] Federal Bureau of Investigation Internet Crime Complaint Center, Senior US officials impersonation campaign. https://www.ic3.gov/PSA/2025/PSA251219

[12] Federal Communications Commission, AI-generated voices in robocalls declaratory ruling. https://docs.fcc.gov/public/attachments/DOC-400393A1.pdf

[13] National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework 1.0. https://www.nist.gov/itl/ai-risk-management-framework

[14] National Institute of Standards and Technology, Generative Artificial Intelligence Profile, NIST AI 600-1. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf

[15] Coalition for Content Provenance and Authenticity, Security Considerations. https://spec.c2pa.org/specifications/specifications/2.4/security/Security_Considerations.html

[16] European Union, Regulation EU 2024 1689 laying down harmonised rules on artificial intelligence. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

[17] European Union, Regulation EU 2022 2065 on a Single Market for Digital Services. https://eur-lex.europa.eu/eli/reg/2022/2065/oj

[18] US Federal Trade Commission, Impersonation of Government and Businesses Rule. https://www.ftc.gov/legal-library/browse/rules/impersonation-government-businesses-rule

[19] IFRS Foundation, IFRS 3 Business Combinations. https://www.ifrs.org/issued-standards/list-of-standards/ifrs-3-business-combinations/

[20] IFRS Foundation, IAS 38 Intangible Assets. https://www.ifrs.org/issued-standards/list-of-standards/ias-38-intangible-assets/

[21] IFRS Foundation, IFRS 13 Fair Value Measurement. https://www.ifrs.org/issued-standards/list-of-standards/ifrs-13-fair-value-measurement/

[22] International Valuation Standards Council, International Valuation Standards. https://ivsc.org/standards/

[23] IFRS Foundation, IAS 36 Impairment of Assets. https://www.ifrs.org/issued-standards/list-of-standards/ias-36-impairment-of-assets/

[24] International Organization for Standardization, ISO IEC 42001 Artificial intelligence management systems. https://www.iso.org/standard/42001

[25] International Organization for Standardization, ISO IEC 23894 Artificial intelligence risk management. https://www.iso.org/standard/77304.html

[26] International Organization for Standardization, ISO IEC 30107 biometric presentation attack detection. https://www.iso.org/standard/67381.html

[27] International Organization for Standardization, ISO IEC 27001 information security management systems. https://www.iso.org/standard/27001

[28] National Institute of Standards and Technology, Digital Identity Guidelines. https://pages.nist.gov/800-63-4/

[29] National Institute of Standards and Technology, Face Analysis Technology Evaluation. https://www.nist.gov/programs-projects/face-analysis-technology-evaluation-fate

[30] National Institute of Standards and Technology, OpenMFC evaluation plan. https://mig.nist.gov/MFC/Web/EvalPlan/OpenMFC2022EvaluationPlan.pdf

[31] European Commission, Guidelines on Transparency of AI-generated Content. https://digital-strategy.ec.europa.eu/en/library/guidelines-transparency-ai-generated-content

[32] European Commission, Technical studies on marking and detecting AI-generated content. https://digital-strategy.ec.europa.eu/en/library/three-studies-technical-solutions-mark-and-detect-ai-generated-content

[33] UK Information Commissioner's Office, Guidance on AI and data protection. https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/

[34] UK National Cyber Security Centre, Guidelines for secure AI system development. https://www.ncsc.gov.uk/collection/guidelines-secure-ai-system-development

[35] Cybersecurity and Infrastructure Security Agency, Secure by Design. https://www.cisa.gov/securebydesign

[36] National Institute of Standards and Technology, Secure Software Development Framework. https://csrc.nist.gov/pubs/sp/800/218/final

[37] National Institute of Standards and Technology, Computer Security Incident Handling Guide. https://csrc.nist.gov/pubs/sp/800/61/r3/final

[38] National Institute of Standards and Technology, Security and Privacy Controls, SP 800-53 Revision 5. https://csrc.nist.gov/pubs/sp/800/53/r5/upd1/final

[39] National Institute of Standards and Technology, Protecting Controlled Unclassified Information in Nonfederal Systems. https://csrc.nist.gov/pubs/sp/800/171/r3/final

[40] European Union Agency for Cybersecurity, ENISA Threat Landscape 2025. https://www.enisa.europa.eu/publications/enisa-threat-landscape-2025

[41] European Union Agency for Cybersecurity, Artificial Intelligence Cybersecurity Challenges. https://www.enisa.europa.eu/publications/artificial-intelligence-cybersecurity-challenges

[42] Organisation for Economic Co-operation and Development, OECD AI Principles. https://oecd.ai/en/ai-principles

[43] UNESCO, Recommendation on the Ethics of Artificial Intelligence. https://www.unesco.org/en/artificial-intelligence/recommendation-ethics

[44] Council of Europe, Framework Convention on Artificial Intelligence and Human Rights Democracy and the Rule of Law. https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence

[45] US Securities and Exchange Commission, Cybersecurity Risk Management Strategy Governance and Incident Disclosure rule. https://www.sec.gov/files/rules/final/2023/33-11216.pdf

[46] Financial Action Task Force, Digital Identity guidance. https://www.fatf-gafi.org/en/publications/Financialinclusionandnpoissues/Digital-identity-guidance.html

[47] International Telecommunication Union, Deepfake detection challenge. https://www.itu.int/en/ITU-T/AI/Pages/deepfake-challenge.aspx

[48] World Intellectual Property Organization, WIPO Conversation on Intellectual Property and Artificial Intelligence. https://www.wipo.int/about-ip/en/artificial_intelligence/conversation.html

[49] UAE Cybersecurity Council, National Cybersecurity Strategy. https://csc.gov.ae/en/strategies/

[50] Saudi National Cybersecurity Authority, Essential Cybersecurity Controls. https://nca.gov.sa/en/regulatory-documents/controls-list/ecc/

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.