American Journal of Computer Science and Technology
2026, Vol. 9, No. 2, pp. 71–83
https://doi.org/10.11648/j.ajcst.20260902.14

SciencePG
Science Publishing Group

Research Article

Architecting a Cryptographically Secure, AI-augmented Paradigm for High-stakes Educational Assessments

Partha Majumdar*

Department of Computer Science, Kalinga University, Raipur, India

Abstract

High-stakes standardised examinations, foundational to meritocratic educational systems, face a crisis of integrity due to systemic vulnerabilities exploited by sophisticated, organised networks. Traditional security measures have proven inadequate against upstream breaches in the physical supply chain, leading to large-scale paper leaks, eroding public trust, and necessitating costly re-examinations. This analysis presents a comprehensive architectural framework designed to fundamentally redesign the assessment ecosystem by integrating advanced, multidisciplinary technologies. The proposed paradigm leverages a permissioned Proof-of-Authority blockchain to create an immutable, transparent ledger for the entire examination lifecycle. Question sourcing is decentralised through a time-distributed micro-sourcing model, in which a vast network of educators submits limited batches of questions, thereby diluting the impact of any single insider threat. Each question is tokenised as Non-Fungible Content (NFC) and stored decentrally, with its provenance secured cryptographically. To achieve perpetual unpredictability and render "guess papers" obsolete, a neuro-symbolic artificial intelligence framework procedurally generates an infinite number of unique, mathematically guaranteed solvable problems. These items are then calibrated using Item Response Theory (IRT), enabling an Automated Test Assembly engine to create millions of individualised yet psychometrically equivalent question papers. The logistical vulnerability of physical transit is neutralised by a secure hybrid edge-printing model, in which encrypted test files are transmitted to examination centres and printed locally just minutes before the test begins, governed by a blockchain-based time lock. By replacing centralised points of failure with cryptographically verifiable, decentralised systems, this framework ensures absolute traceability, psychometric equity, and operational resilience, thereby restoring the sanctity and credibility of high-stakes educational assessments.

Keywords

High-stakes Exam Integrity, Blockchain and NFC Tokenisation, Neuro-symbolic AI, Item Response Theory (IRT), Secure Hybrid Edge-printing

1. Introduction

The administration of high-stakes standardised testing forms the foundational bedrock of meritocratic educational systems and professional credentialing frameworks globally. These examinations determine the trajectory of millions of lives, allocating scarce institutional resources based on objective, quantifiable metrics of academic proficiency. However, escalating complexities in the logistical execution of large-scale, pen-and-paper examinations have exposed critical

*Correspondence: Partha Majumdar (partha.majumdar@hotmail.com)

Received: 15 May 2026; **Accepted:** 25 May 2026; **Published:** 12 June 2026

structural vulnerabilities that threaten the very concept of meritocracy. In India, the National Eligibility cum Entrance Test (NEET-UG) represents one of the largest single-day, single-shift examinations in the world, with over 2.4 million candidates competing fiercely for approximately 100,000 undergraduate medical seats. The intense competition, coupled with archaic logistical frameworks and immense sociological pressure, has inadvertently spawned sophisticated, highly resourced "education mafias" that exploit systemic weaknesses to orchestrate massive breaches of examination integrity.

The ramifications of these breaches extend far beyond immediate administrative chaos; they fundamentally erode public trust, precipitate widespread anxiety among student populations, and trigger exhaustive judicial, political, and federal interventions. The traditional, reactive responses to these breaches—such as intensified physical frisking, the deployment of signal jammers, and localised biometric verification at examination centres—have proven repeatedly insufficient. The locus of vulnerability has decisively shifted upstream, from localised examination hall malpractice to the pre-examination supply chain, where encrypted digital messaging platforms facilitate the rapid, untraceable dissemination of compromised materials nationwide.

To restore the sanctity of high-stakes evaluations, a radical, evolutionary leap in examination architecture is required. This report presents an exhaustive, multidisciplinary framework that fundamentally redesigns the assessment ecosystem from question conception to the final mile of delivery. By synthesising distributed ledger technology, Non-Fungible Content tokenisation, advanced neuro-symbolic artificial intelligence, Item Response Theory, and secure hybrid edge-printing logistics, this framework is designed to neutralise both internal and external threat vectors. The proposed architecture replaces centralised, trust-based points of failure with cryptographically verifiable, decentralised systems, ensuring absolute traceability, psychometric fairness, and operational resilience.

2. Diagnostic Analysis of Systemic Vulnerabilities in Examination Logistics

To engineer a mathematically and logistically resilient solution, it is first imperative to dissect the mechanical and human failures that plague current assessment models. The recurrent breaches in national examinations are rarely spontaneous anomalies; rather, they are the result of calculated, highly organised exploitation of logistical and procedural choke points that exist within a physically distributed supply chain.

2.1. The Historical Context of Assessment Breaches

The vulnerability of the Indian examination system is not a recent phenomenon but a persistent systemic flaw that has escalated in tandem with advances in digital communication. Historically, the 2015 All India Pre-Medical Test (AIPMT) was completely cancelled by the Supreme Court of India due to a large-scale paper leak, setting a precedent for judicial intervention in academic administration. Subsequent years witnessed parallel controversies, including the 2018 Central Board of Secondary Education (CBSE) leaks, the 2017 Staff Selection Commission (SSC) protests, and the 2021 Joint Entrance Examination (JEE) Main hacking incident, where remote access tools were utilised to bypass computer-based testing protocols.

However, the controversies surrounding the NEET-UG examinations in 2024 and 2026 have brought these vulnerabilities to a critical inflexion point. In 2024, the examination was marred by allegations of localised paper leaks in Bihar and Jharkhand, coupled with widespread outrage over the awarding of grace marks due to administrative delays at specific centres. This resulted in an unprecedented anomaly where 67 candidates secured perfect scores, a statistical impossibility under normal testing conditions. While the Supreme Court ultimately ruled that the 2024 breach was localised rather than systemic, the event severely damaged institutional credibility.

This credibility crisis culminated in the devastating cancellation of the NEET-UG 2026 examination. Investigations by the Rajasthan Special Operations Group (SOG) and the Central Bureau of Investigation (CBI) revealed a highly organised syndicate that had intercepted the examination materials weeks in advance. A handwritten "guess paper" containing over 410 questions was circulated via encrypted messaging platforms such as WhatsApp and Telegram. Forensic analysis confirmed that approximately 120 questions from the chemistry and biology sections of this illicit document matched the official examination paper exactly. The government was forced to order a complete annulment of the examination for 2.2 million students, recognising that the integrity of the medical merit list had been irreversibly compromised.

2.2. Upstream and Downstream Choke Points

The traditional lifecycle of a standardised question paper involves prolonged exposure to multiple high-risk vectors, broadly categorised into upstream and downstream phases. The upstream process of question setting typically occurs in highly secluded environments, where panels of subject-matter experts are sequestered to generate pools of questions without access to digital tools. While this stage is heavily fortified, it remains inherently dependent on the absolute incorruptibility of a small group of human actors.

The downstream stages, however, represent the most critical vulnerabilities. These involve the transfer of finalised manuscripts to centralised commercial printing presses, the mass production of millions of physical booklets, and the subsequent transportation of these materials via armed transit to

regional strongrooms and local banking vaults across vast geographic expanses. Each node in this physical supply chain—from the printing press operator to the logistics contractor and the local strongroom custodian—introduces a human intermediary susceptible to financial coercion or social engineering. The 2026 breach explicitly demonstrated that sophisticated syndicates do not need to breach the examination hall if they can intercept the material at the printing or transit stage and monetise it digitally. Consequently, the integrity of the entire examination ecosystem becomes constrained not by the sophistication of the assessment design itself, but by the weakest operational link in the physical distribution chain.

2.3. The Legislative Framework and Liability Mandates

Recognising the severity of these organised syndicates, the Indian legislature enacted the *Public Examinations (Prevention of Unfair Means) Act, 2024*. This landmark legislation provides a stringent legal framework to deter examination malpractice, categorising such acts as cognizable, non-bailable, and non-compoundable offences. The Act prescribes rigorous penalties, including imprisonment for up to 10 years and fines of up to ₹1 crore for individuals involved in organised examination crimes.

Crucially, the legislation imposes strict liability on commercial service providers and logistical partners engaged by the examination authorities. If a service provider is found to be complicit in, or negligent regarding, a breach, they face massive financial penalties, the attachment of corporate property, and a multi-year ban from participating in public examination contracts. [9] While the Act establishes a formidable punitive deterrent, legislation alone cannot preempt technical or physical breaches. The law must be operationalised through an infrastructure that provides incontrovertible digital forensics and makes the illicit extraction of data computationally and physically impossible. The technological paradigm proposed in this report provides the exact cryptographic assurance required to comply with and enforce the mandates of this legislative framework.

3. Decentralised Question Sourcing: Mitigating Insider Threats

The conventional method of assembling question panels in centralised, physical locations creates single points of vulnerability and severely limits the diversity and volume of the question pool. A mathematically superior approach involves democratising question generation through secure, distributed networks, with asynchronous submission protocols to mitigate the risk of compromise by a question provider.

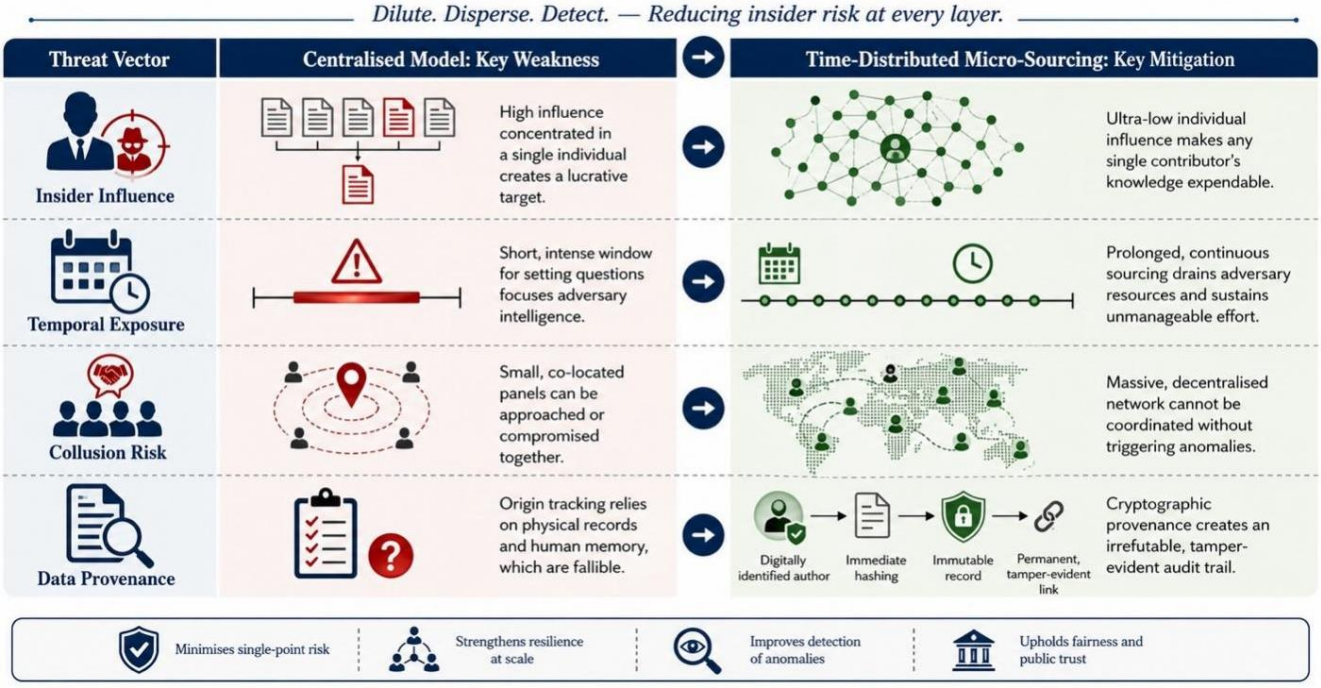

Figure 1. *Strengthening Assessment Integrity Through Distributed Sourcing.*

The Time-Distributed Micro-Sourcing Model

Insider threats are consistently identified as critical risks in data security and examination logistics, often motivated by financial gain, discontent, or coercion from external syndicates. In a traditional setup, if a senior examiner who has contributed 20% of the final paper is compromised, the entire examination is rendered

invalid. To completely neutralise this vector, the architecture must systematically limit the exposure and influence any single contributor has over the final examination compilation.

The proposed solution implements a rigorous *Time-Distributed Micro-Sourcing protocol*. Under this model, the examination authority curates an expansive, geographically distributed network of vetted educators, academics, and subject-matter experts. Instead of convening for a brief, intensive period, these contributors are required to submit a strictly limited number of questions—for example, a maximum of five items per educator—over a prolonged, asynchronous, chronological window spanning six to eight months.

This model introduces two powerful mathematical defences against insider leaks:

First, it achieves a radical dilution of influence. If an individual contributor decides to act maliciously and leaks their allotted five questions to a coaching syndicate, the statistical probability of those specific items appearing in a randomly assembled, 180-question examination drawn from a master repository of tens of thousands of items is mathematically negligible. The leaked questions simply vanish into the statistical noise of the larger pool.

Second, it provides mitigation against systemic network breaches. For an organised syndicate to assemble a viable "guess paper" that guarantees a high score (as witnessed in the 2026 controversy), they would need to successfully compromise, coordinate, and collude with a massive, geographically disparate proportion of the independent question setters over an extended period. This dramatically increases the logistical complexity and the risk of detection for the syndicate, making a coordinated upstream breach practically impossible.

4. Cryptographic Provenance: Blockchain and Non-fungible Content

Upon submission of these micro-batches of questions, the data must be digitised and secured to definitively prevent alteration, unauthorised deletion, or untracked duplication. Relying on traditional centralised databases (such as standard SQL servers) merely shifts the point of failure from a physical strongroom to a digital server, remaining vulnerable to database administrators or sophisticated cyber intrusions. [1] The solution lies in implementing distributed ledger technology and tokenising examination items as Non-Fungible Content.

4.1. Non-fungible Content (NFC) Tokenisation

Every individual question submitted to the central repository is mathematically classified and minted as "Non-Fungible Content" (NFC). While the broader public is familiar with Non-Fungible Tokens (NFTs) in the context of digital art and speculative assets, the underlying cryptographic standards (such as ERC-721 or ERC-1155 on Ethereum-compatible networks) are designed to prove the absolute authenticity, provenance, and unique identity of any digital asset.

When an educator uploads a question, the system applies a cryptographic hash function (such as SHA-256) to the digital file. This generates a fixed-length string of characters—a unique digital fingerprint that changes entirely if even a single pixel or character of the source file is altered. [16] This hash, along with metadata regarding the author, subject, and timestamp, is minted as an NFC on the blockchain. [22] This ensures that the question becomes an indelible part of the educational ledger, with its creation and access history permanently preserved and immune to internal database manipulation.

4.2. Decentralised Storage via IPFS

Storing millions of high-resolution images, complex mathematical formulas, and text files directly on a blockchain is computationally inefficient and prohibitively expensive. Therefore, the architecture separates the immutable ledger from physical data storage using the InterPlanetary File System (IPFS). [20].

IPFS is a peer-to-peer distributed file storage protocol that fundamentally alters how data is retrieved. Instead of using location-based addressing (e.g., fetching a file from a specific centralised server URL), IPFS utilises content-based addressing. The heavily encrypted raw file of the question is uploaded to the IPFS network, which fragments the data and distributes it across multiple decentralised nodes. [20] The network returns a unique Content Identifier (CID) derived directly from the file's cryptographic hash. This CID is stored in the blockchain's NFC smart contract. [22] This dual-layer architecture guarantees that the actual content remains encrypted and decentralised, eliminating the vulnerability of a central server honeypot, while the blockchain ledger maintains an unalterable audit trail of its existence. [26].

4.3. Proof of Authority Consensus and Smart Contract Governance

Public, permissionless blockchains (such as the Bitcoin or Ethereum mainnets) operate on Proof of Work (PoW) or Proof of Stake (PoS) consensus mechanisms. These networks prioritise absolute, trustless decentralisation but suffer from highly variable transaction speeds, exorbitant computational and financial costs, and a lack of regulatory control.

For sovereign national educational boards, a permissioned blockchain utilising a Proof of Authority (PoA) consensus algorithm provides the optimal governance structure. In a PoA network, the mathematical consensus required to validate blocks of transactions is achieved not through competitive computational mining or cryptocurrency staking, but by relying on a select group of pre-approved, legally accountable validator nodes. These nodes would be operated by highly trusted institutional entities, such as the vice-chancellors of premier central universities, directors of the Indian Institutes of Technology (IITs), and the leadership of the National Informatics

Centre (NIC).

In this structure, validators stake their legal identities and institutional reputations rather than digital capital. Any anomalous or malicious action on the network is immediately attributable to a known, sovereign entity, aligning perfectly with the rigorous punitive frameworks established by the 2024 Unfair Means Act. [9] Furthermore, PoA networks offer exceptionally high-speed transaction throughput and negligible energy consumption, which is essential for processing millions of concurrent student records and item bank interactions during peak examination windows. [18].

The interaction with the IPFS-stored question bank is strictly governed by self-executing smart contracts written in languages such as Solidity. [20] These contracts automate Role-Based Access Control (RBAC) and implement irrevocable Time-Locked Encryption. A smart contract can be hard-coded to reject any decryption-key request from the network until a precisely synchronised global timestamp is reached (for instance, exactly 10 minutes prior to the physical commencement of the examination). Because the rules are embedded in the blockchain's consensus layer, this time-lock cannot be overridden by any single administrator, IT technician, or government official, rendering pre-examination digital leaks technically impossible. [22].

To further protect the physical safety of the question setters and insulate them from targeted coercion by external syndicates, Zero-Knowledge Proofs (ZKPs) are integrated into the authentication layer. ZKPs are cryptographic protocols that allow a prover to convince a verifier that a statement is true without revealing any information beyond the statement's absolute validity. [11] In this context, a ZKP allows the system to definitively verify that a submitted question originated from an authorised, credentialed educator with the correct cryptographic keys, without ever recording or revealing the educator's specific identity, location, or name to the broader network or database administrators. [26].

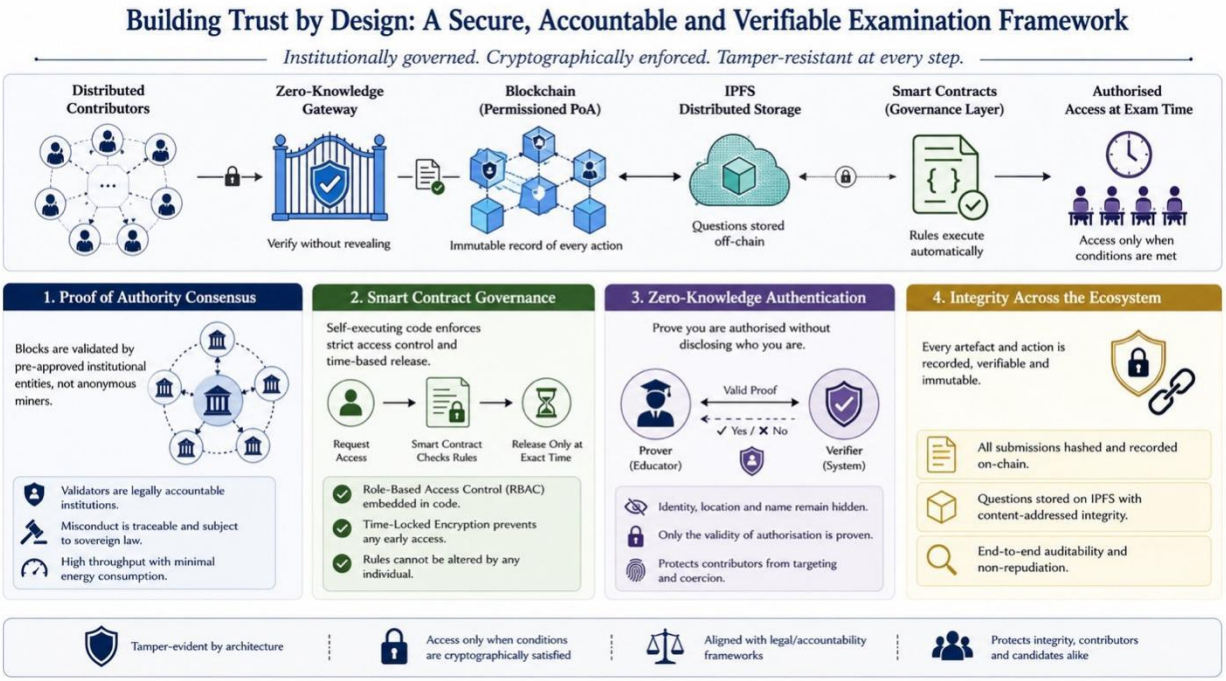

Figure 2. *Building Trust by Design.*

5. Neuro-symbolic Artificial Intelligence in Procedural Item Generation

While time-distributed micro-sourcing, protected by zero-knowledge proofs, ensures a vast and cryptographically secure repository of baseline academic concepts, human-generated item banks are inherently finite. Over time, even massive databases become susceptible to eventual mapping and memorisation by highly resourced coaching syndicates that deploy thousands of students to memorise and reconstruct past papers. To achieve true, perpetual unpredictability, the assessment architecture must evolve beyond static storage and employ Artificial Intelligence to dynamically and procedurally generate scientific, quantitative questions.

5.1. Overcoming Hallucinations in Pure Generative Models

Large Language Models (LLMs), such as advanced itera-

tions of GPT, Claude, and Gemini, have demonstrated profound capabilities in natural language processing, semantic reasoning, and even achieving high percentiles on standard medical and legal examinations. [6] However, relying solely on pure generative LLMs for the autonomous creation of high-stakes physics, chemistry, or complex mathematical problems presents an unacceptable operational risk.

Pure LLMs are fundamentally probabilistic engines; they generate responses based on statistical word associations rather than an intrinsic understanding of physical laws or deterministic mathematics. [2] Consequently, when tasked with generating novel quantitative problems, they are highly prone to "hallucinations"—producing questions that appear linguistically coherent but are mathematically intractable, conceptually flawed, or lack feasible, unambiguous solutions. [6] In a high-stakes environment like NEET, where a single erroneous question can trigger mass litigation, mandate the awarding of grace marks, and disrupt the entire national admission cycle, stochastic generation without deterministic verification is entirely unviable.

5.2. The Implementation of Neuro-symbolic Frameworks

The solution to achieving infinite, reliable scalability lies in deploying Symbolic AI and Neuro-Symbolic frameworks. Symbolic AI operates on high-level, human-readable representations of logic, rigid rules, and deterministic mathematics. By integrating the fluid, context-generating capabilities of neural networks with the rigorous, rule-bound engines of symbolic mathematics, the system can generate an infinite number of procedural variations of a baseline problem that are mathematically guaranteed to be solvable. [13].

A prominent structural example of this architecture is the Symbolic Integration for Generative Systems (SIGS) framework, which utilises formal grammars and deterministic solvers to ensure that generated equations and physical scenarios are syntactically and physically valid by construction. [13] When this neuro-symbolic approach is applied to the examination pipeline, the process operates as follows:

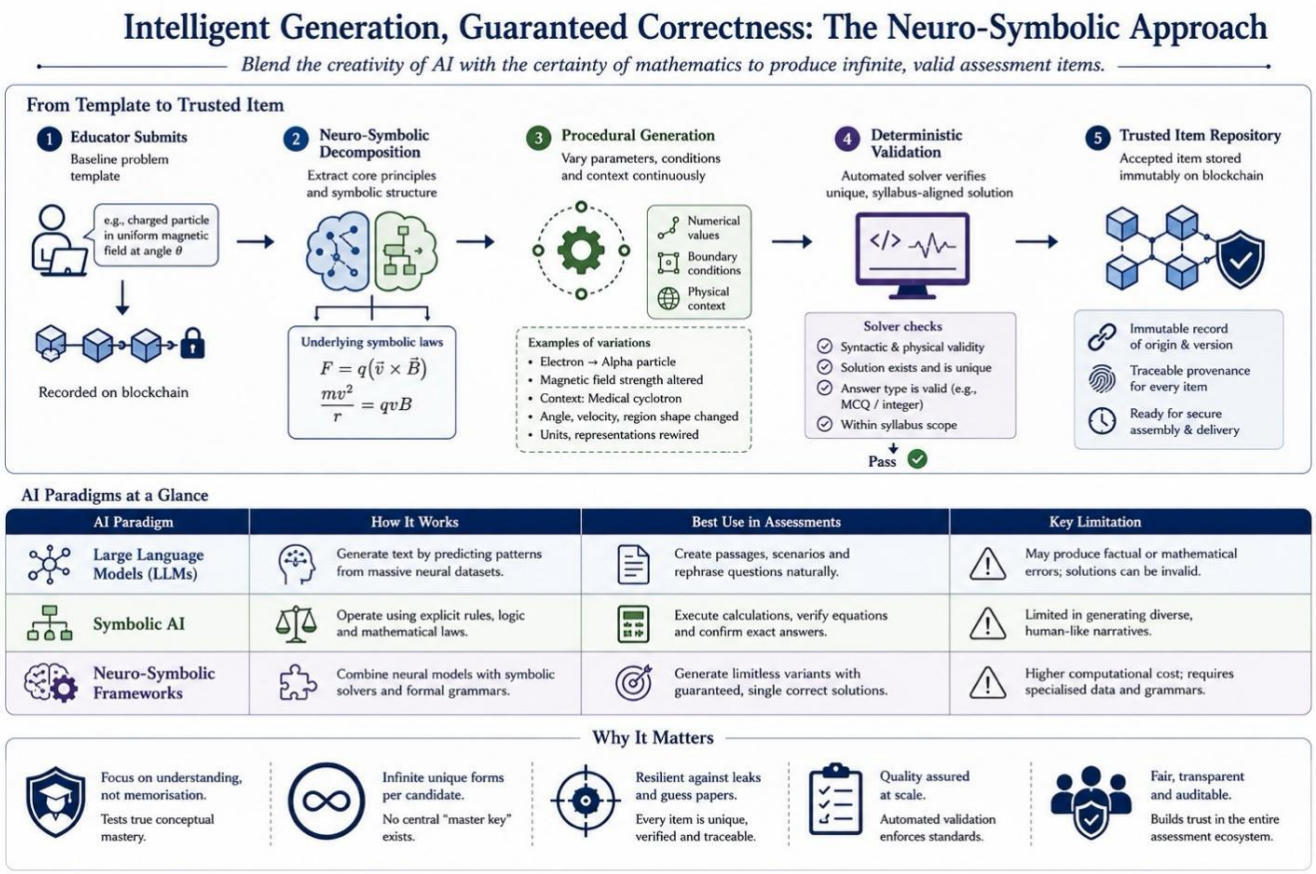

Figure 3. *Intelligent Generation, Guaranteed Correctness.*

First, an authorised educator submits a baseline problem template to the blockchain (for example, calculating the trajectory of a charged particle entering a uniform magnetic field at a specific angle). Second, the AI deconstructs this template into its core physical principles, isolating the underlying symbolic equations (e.g., the Lorentz force equation and centripetal kinematics) from the descriptive text. Third, the procedural generation engine continuously alters the numerical parameters, the

boundary conditions, and the contextual framing of the problem. It might change the particle from an electron to an alpha particle, alter the magnetic field strength, or rewrite the narrative context to describe a medical cyclotron rather than a generic laboratory vacuum.

Crucially, before any generated item is accepted into the active blockchain repository, it is passed through an integrated automated solver—a computational engine akin to Wolfram Alpha or a specialised, fine-tuned physics AI solver. [10] The deterministic solver executes the problem using the newly generated parameters to confirm that they yield a definitive, unambiguous integer or multiple-choice solution that falls strictly within the prescribed syllabus scope.

This neuro-symbolic procedural generation ensures that candidates are tested on genuine conceptual mastery and analytical reasoning, rather than their ability to memorise specific numerical combinations provided by illicit "guess papers" circulated by coaching mafias. [7] Because the exact numerical values and framing of the question are generated procedurally and mapped directly to the individual candidate's unique test form, the concept of a master "answer key" ceases to exist.

5.3. Human-in-the-Loop (HITL) Vigilance and Bias Mitigation

Despite the mathematical rigour provided by neuro-symbolic solvers, algorithmic generation must remain subject to Human-in-the-Loop (HITL) vigilance. Automated systems, while mathematically flawless, can inadvertently generate descriptive scenarios that are practically absurd or contain implicit cultural or socioeconomic biases that could disadvantage specific student demographics. [3].

Therefore, a secondary layer of blinded, decentralised expert reviewers—operating through the same PoA blockchain interface—must validate a randomised sample of the procedurally generated questions. This review ensures pedagogical clarity, appropriate reading levels, and construct validity before the items are permanently hashed and stored in the IPFS repository. [26] This symbiotic relationship between AI generation and human oversight provides infinite scalability while maintaining absolute educational integrity.

6. Psychometric Fairness in Individualised Assessments

A central pillar of the proposed solution is the generation of truly individualised question papers for each candidate, thereby rendering mass cheating, paper leaks, and organised solver syndicates fundamentally obsolete. If every candidate in an examination hall, or across the country, receives a uniquely compiled paper containing procedurally generated questions, copying from a neighbour or purchasing a leaked answer key becomes structurally useless.

However, the transition from a single national paper to individualised assessments poses a significant challenge to psychometric fairness. Pure, unweighted randomisation of questions poses a severe threat to equitable evaluation. If Candidate A randomly receives a disproportionate number of highly complex, multi-step physics questions while Candidate B receives predominantly foundational, formula-based items, the resulting scores are fundamentally incomparable. In a high-stakes environment where percentiles dictate medical admissions, such statistical disparities violate the core principles of equity, undermine the validity of the test, and would inevitably invite massive, paralysing legal challenges. [14].

Furthermore, fairness in large-scale assessments is not merely a pedagogical concern but also a statistical and constitutional necessity. Any perception that one candidate received an easier or more advantageous examination form can rapidly erode public trust in the credibility of the entire admission process. This challenge becomes even more pronounced in decentralised or AI-generated assessment environments, where millions of unique item combinations may exist simultaneously. Consequently, the examination architecture must incorporate rigorous psychometric calibration mechanisms to ensure that every candidate, regardless of the questions they receive, encounters an examination with mathematically equivalent difficulty, discrimination, and scoring reliability.

6.1. Overcoming the Limits of Classical Test Theory with IRT

To facilitate individualised examinations without compromising perceptions of fairness or its reality, the system must abandon outdated evaluation models and fully integrate Item Response Theory (IRT). Classical Test Theory (CTT), the traditional framework used for decades, evaluates a test based entirely on the total raw score, if all items contribute equally to the measurement of the candidate's proficiency. [4] CTT metrics are entirely sample-dependent; a question appears "hard" only if the specific group of students taking the test performs poorly on it, making it impossible to accurately compare scores across different groups taking different sets of questions. [8].

Item Response Theory, conversely, models the probabilistic relationship between a candidate's underlying, latent ability (denoted in psychometrics as theta, θ) and the distinct statistical properties of individual test items. [4] Under advanced IRT frameworks, such as the 3-Parameter Logistic (3PL) model, every single question in the decentralised IPFS bank is rigorously calibrated through pilot testing and historical data analysis to determine three critical, stable metrics:

First, the Item Difficulty (b-parameter) specifies the exact point on the capability scale at which a candidate has a 50% probability of answering the question correctly. Second, the Item Discrimination (a-parameter) measures the slope of the item characteristic curve, indicating how effectively the item differentiates between high- and low-ability candidates. [8]

Third, the Pseudo-Guessing (c-parameter) accounts for the probability that a candidate with extremely low ability will answer the question correctly simply by guessing, a factor that is mathematically crucial for multiple-choice formats like NEET. [4].

Recent advancements in computational psychometrics, particularly the emergence of Fair-IRT frameworks, extend significantly beyond the traditional objectives of basic item calibration and score normalisation. These advanced models systematically analyse Item Characteristic Curves (ICCs) to determine whether specific questions exhibit Differential Item Functioning (DIF), a condition in which candidates from different demographic or educational backgrounds show unequal probabilities of answering an item correctly despite having equivalent underlying ability levels. [8] Such disparities may arise from hidden linguistic assumptions, culturally specific references, unequal access to educational resources, or regional variations in pedagogy, thereby introducing unintended bias into the assessment process.

By continuously evaluating these statistical patterns across large and diverse candidate populations, Fair-IRT systems can detect subtle inequities related to gender, socioeconomic status, language background, geographic region, or prior educational exposure. [25] Questions identified as exhibiting significant DIF can then be revised, recalibrated, or permanently removed from the active item repository. This process ensures that examination outcomes are driven by genuine conceptual understanding and analytical competence rather than by demographic privilege or contextual familiarity. Consequently, advanced IRT-based fairness auditing strengthens the psychometric validity, transparency, and social legitimacy of large-scale national examinations. [12].

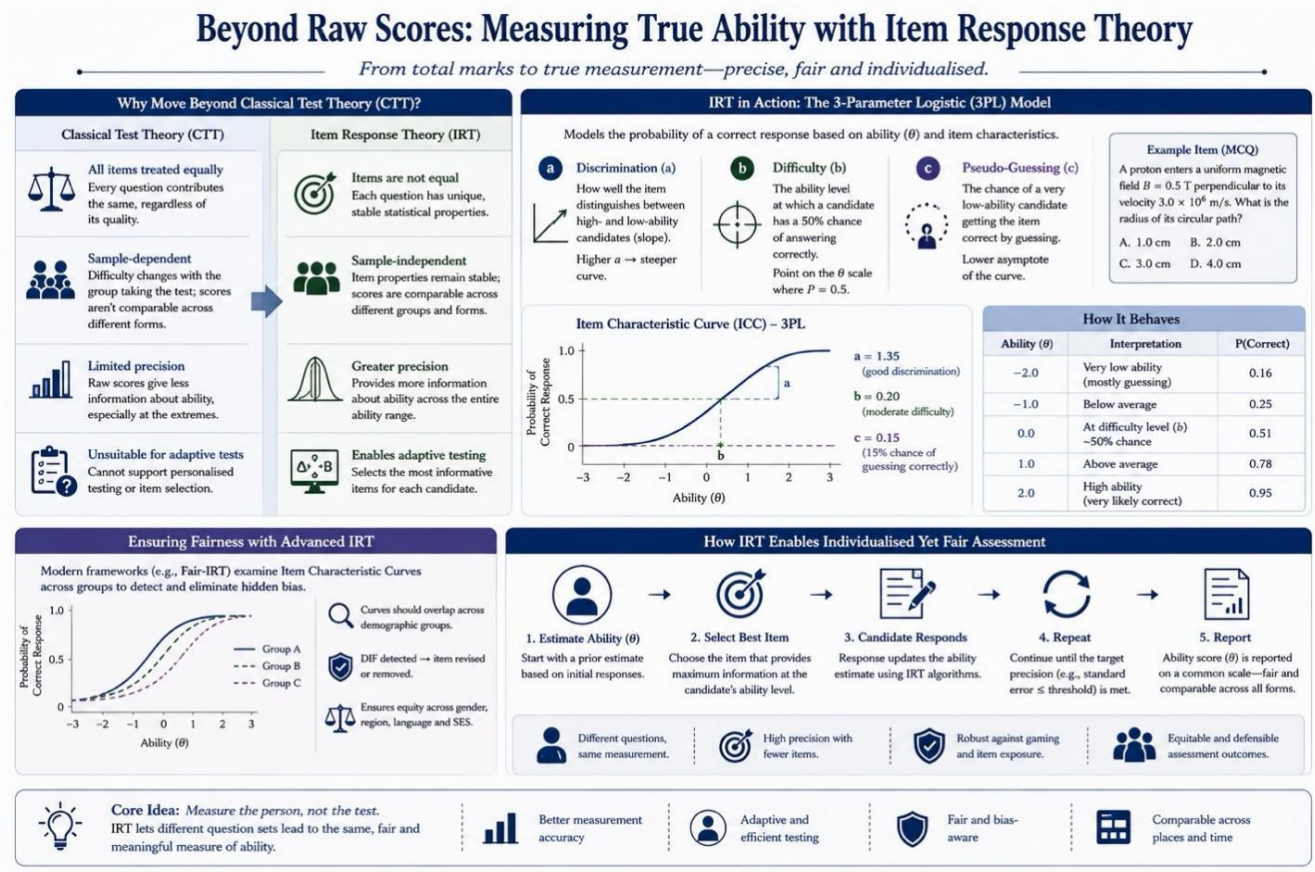

Figure 4. *Building Raw Scores.*

6.2. Automated Test Assembly (ATA) Using Linear Programming

Once the item bank is calibrated via IRT, Automated Test Assembly (ATA) replaces manual question paper compilation. ATA uses operations research, specifically Mixed-Integer Linear Programming (MILP), to efficiently select questions meeting strict constraints from the vast item bank.

When the ATA engine is tasked with assembling an individualised paper for a candidate, it applies a matrix of concurrent constraints:

1) *Content and Categorical Constraints:* The algorithm ensures exact alignment with the national syllabus blueprint. For instance, it uses linear constraints to ensure that each unique paper contains precisely 45 Chemistry

questions, with exactly 15 from Organic Chemistry, 15 from Inorganic Chemistry, and 15 from Physical Chemistry, matching the prescribed topical weighting of the medical commission. [5].

2) *Target Information Functions (TIF):* Utilising complex MAXIMIN and MINIMAX objective functions, the ATA engine ensures that the aggregated difficulty and discrimination curves of the selected items perfectly match a universally predefined target difficulty curve for the overall examination. [5].

Consequently, while two candidates sitting next to each other may receive completely different questions featuring procedurally generated numerical values, the psychometric weight, the topical distribution, and the statistical difficulty of their respective examinations are mathematically identical.

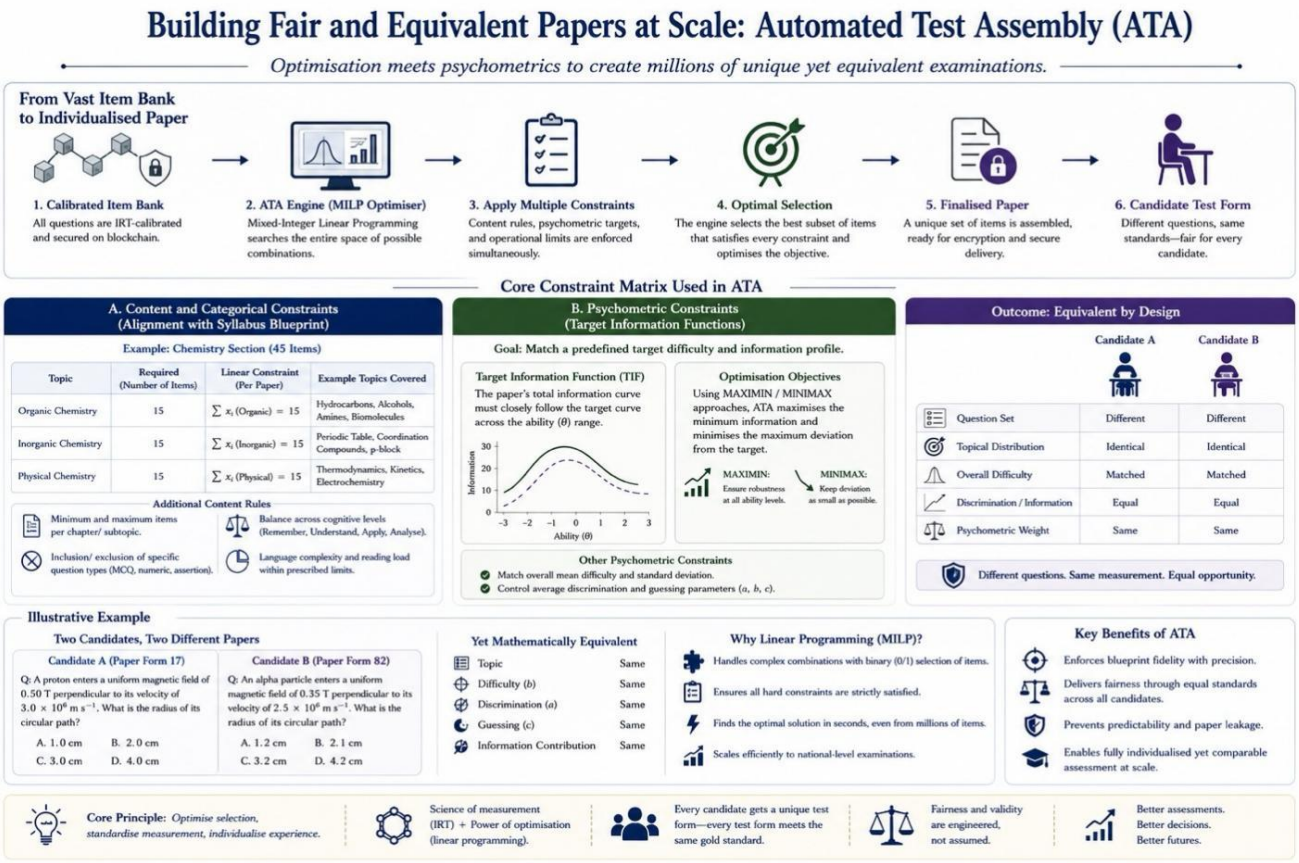

Figure 5. *Building Fair and Equivalent Papers at Scale.*

6.3. Score Equating and Normalisation Protocols

Even with the most robust ATA algorithms, microscopic, statistically significant differences in the overall difficulty of individualised forms are inevitable at the scale of millions of candidates. To ensure absolute parity before issuing national rankings, post-examination Score Equating must be applied. Equating is a rigorous statistical procedure that adjusts raw scores to compensate for residual differences in difficulty, placing all candidates' performances on a common, universally comparable scale. [15].

By deliberately embedding a small set of highly secure, pre-calibrated common "anchor items" within the randomised sets, the system can utilise IRT true-score equating methodologies to generate normalised percentiles. [21] This ensures that an equated score of 650 on one unique test form represents the exact same level of medical proficiency as a score of 650 on any other unique form. This comprehensive psychometric framework resolves the primary judicial and logistical hurdles that currently prevent the adoption of randomised testing in India, ensuring that the examination remains a pure, legally defensible measure of merit.

7. Edge Delivery Logistics: The Secure Hybrid Model

The ultimate logistical challenge in this decentralised paradigm lies in the physical delivery of the securely assembled, individualised examinations to the candidates on the day of

the test. A fully online, Computer-Based Test (CBT) administered simultaneously to 2.4 million users is currently constrained by severe infrastructural deficits across the subcontinent. Providing uninterrupted, high-bandwidth internet connectivity, secure computer terminals, and highly stable power grids to thousands of examination centres in rural and semi-urban India in a single day is logistically unfeasible.

Attempting to circumvent this by transitioning to multiple CBT shifts spread over several weeks introduces significant complexities in normalisation, extends the agonizing, high-stress period for students, and significantly increases the temporal window for cyber-attacks and server manipulation. The optimal solution, in line with the strategic recommendations of the government-appointed Radhakrishnan Committee, is to implement a Computer-Assisted Secure Paper-Based Test (Hybrid Model). This hybrid architecture strategically combines the scalability and cryptographic security of digital systems with the operational familiarity, accessibility, and resilience of conventional paper-based examinations.

7.1. Encrypted Transmission and Edge Printing

This hybrid system eliminates the vulnerability of physically transporting printed booklets. The ATA engine first finalises question papers on a secure blockchain. Then, hours or minutes before exams, these encrypted digital packages are sent via secure VPNs to local servers within examination centres or regional hubs.

Third, and most critically, these encrypted files cannot be opened upon receipt. They remain dormant until the blockchain smart contract's time-lock mechanism is activated. At the precise, universally synchronised moment, the decryption keys are released to the centre administrators, who must provide multi-factor authentication (MFA) and biometric sign-off to access them. [11].

Once decrypted locally, files go to secure, high-speed digital printers in a strongroom. These printers produce 50-100 pages per minute. A centre with 500 students can print all papers in under an hour. The software assigns a unique serial number to each booklet. Each page has a barcode, a QR code, and a watermark with candidate details that link it to their digital identity.

This "just-in-time" methodology shrinks the massive vulnerability window from weeks of cross-country transit and exposure to commercial printing presses to mere minutes of supervised, localised printing, effectively neutralising the threat of interception in transit and large-scale physical leaks. [23].

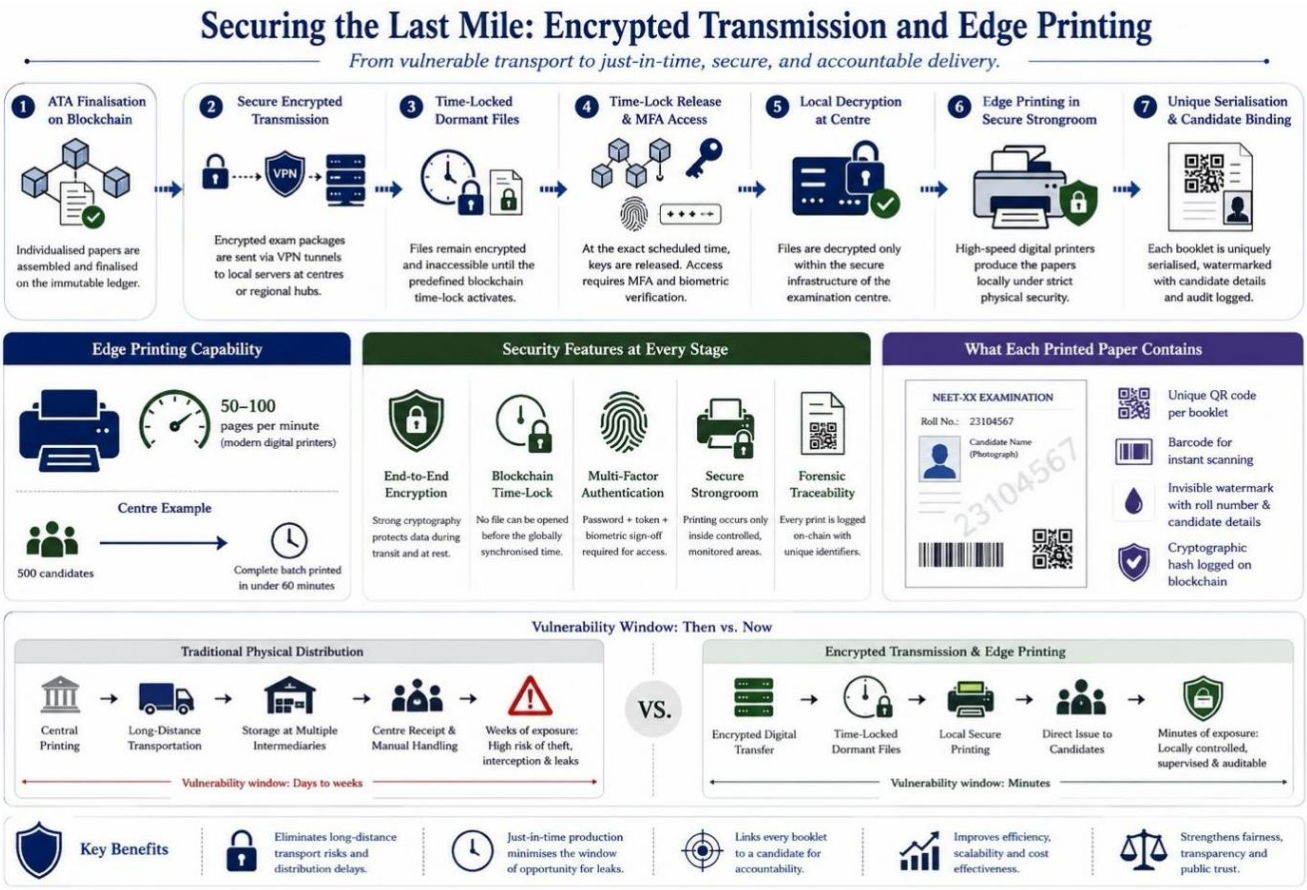

Figure 6. *Securing the Last Mile.*

7.2. Physical Security via NFC Tracking

While the digital supply chain is now cryptographically secured, the final physical handling of the freshly printed papers must still be strictly monitored to prevent localised centre-level corruption. To ensure an unbroken chain of custody within the exam centre, the sealed envelopes containing the individualised papers are equipped with physical Near Field Communication (NFC) tags. [24].

These tamper-evident, cryptographic NFC tags offer rigorous "Proof of Presence" tracking. As the envelopes are moved from the secure printing room to the specific classrooms, authorised invigilators must tap the tags using specialised, institution-provided mobile devices. Each tap generates a unique, encrypted, timestamped, and geo-located URL that is instantly logged on the PoA blockchain, verifying exactly who handled the package, when, and in which specific corridor of the school. [24] Furthermore, if the envelope's physical seal is broken prematurely, the NFC tag's internal antenna is severed, or its status changes, it immediately triggers a system-wide alert to the central command hub. This allows authorities to instantly isolate a breach to a specific room or invigilator, rather than compromising the entire national examination grid.

8. Economic Feasibility, Scalability, and Legal Compliance

A fundamental and often-cited critique of deploying advanced technologies in developing nations is the capital expenditure required. Deploying commercial-grade, high-speed digital printers, localised secure servers, and NFC-based tracking infrastructure across thousands of examination centres represents a significant initial investment in infrastructure. [17].

However, this capital expenditure must be analysed and contextualised against the recurring, astronomical economic and social costs of the current failing system. The financial burden of chartering fleets of physical armed transport, executing massive central printing contracts, paying for prolonged strongroom guarding, and the catastrophic economic losses associated with cancelling, refunding, and re-conducting national examinations for millions of students far outweighs the amortisation of decentralised digital hardware over a multi-year period. Furthermore, examination bodies have substantial financial capacity; reports indicate that entities like the NTA generate significant annual surpluses (e.g., a ₹448 crore surplus between 2018 and 2024), providing the internal capital needed to finance this infrastructural transition without imposing additional fee burdens on candidates. [19].

Crucially, this comprehensive architecture ensures strict compliance with the mandates of the *Public Examinations (Prevention of Unfair Means) Act, 2024*. By recording every single digital and physical interaction on an immutable PoA blockchain—from the identity of the AI-augmented question setter to the local printer operator and the classroom invigilator—the system establishes incontrovertible, automated digital forensic evidence. [9] If an anomaly or a localised attempt at malpractice occurs, federal investigative agencies like the CBI no longer need to rely on prolonged, opaque interrogations or easily destroyed paper trails. The immutable ledger instantly identifies the precise node, the exact timestamp, and the specific authorised individual responsible for the breach, enabling swift, targeted, and legally airtight prosecution.

9. Conclusion

The persistent, escalating compromises of India's high-stakes medical and engineering entrance examinations signify the terminal decline of legacy, centralised assessment logistics. Attempting to secure a massive, paper-based examination through localised physical policing and reactive legislation is fundamentally misaligned with the realities of modern, digitally enabled organised crime networks.

The seamless integration of blockchain-secured micro-sourcing, neuro-symbolic AI procedural generation, IRT-driven automated test assembly, and secure hybrid edge-printing logistics represents a comprehensive, much-needed paradigm shift. This architecture mathematically isolates and neutralises insider threats, renders the mass memorisation of leaked materials entirely obsolete through individualised, psychometrically-equated test forms, and collapses the physical vulnerability window from weeks of transit to mere minutes of secure, localised production.

Implementing this decentralised framework undoubtedly demands decisive administrative will, robust pilot testing, and substantial capital investment in digital infrastructure at the examination centre level. However, the ultimate dividend is the restoration of absolute, unassailable credibility to the national examination process. By transitioning from an archaic system reliant on fragile human trust to one governed by immutable cryptographic truth, national testing agencies can finally fulfil their primary mandate and guarantee what every student fundamentally deserves: an impregnable, fair, and purely meritocratic evaluation.

Abbreviations

3PL	3-Parameter Logistics
AII	Artificial Intelligence
AIPMT	All India Pre-Medical Test
ATA	Automated Test Assembly
CBI	Central Bureau of Investigation
CBSE	Central Board of Secondary Education
CBT	Computer-Based Test
CID	Content Identifier
CTT	Classical Test Theory
GPT	General Purpose Transformer

HITL	Human-in-the-Loop
ICC	Item Characteristic Curves
IPFS	Inter-Planetary File System
IRT	Item Response Theory
JEE	Joint Entrance Examination
LLM	Large Language Model
MFA	Multi-Factor Authentication
MILP	Mixed-Integer Linear Programming
NEET-UG	National Eligibility cum Entrance Test- Under Graduate
NFC	Near Field Communication
NFC	Non-Fungible Content
NFT	Non-Fungible Token
NIC	National Information Centre
NTA	National Testing Agency
PoA	Proof of Authority
PoS	Proof of Stake
PoW	Proof of Work
QR	Quick Response
SIGS	Symbolic Integration of Generative Systems
SOG	Special Operations Group
SQL	Structured Query Language
SSC	Staff Selection Commission
TIF	Target Information Functions
URL	Universal Resource Locator
ZKP	Zero-Knowledge Proof

Author Contributions

Partha Majumdar: Conceptualization, Formal Analysis, Methodology, Project Administration, Resources, Validation, Visualization, Writing – original draft, Writing – review & editing

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] Abdelsalam, M., Shokry, M., & Idrees, A. M. (2024, January 18). A proposed model for improving the reliability of online exam results using blockchain. *IEEE Access*, *12, 7719-7733.* https://doi.org/10.1109/ACCESS.2023.3304995

[2] Bralin, A., & Rebello, N. S. (2025). AI reasoning models for problem solving in physics. *arXiv*. https://arxiv.org/html/2508.20941v1

[3] Burke, C. M. (2025). AI-assisted exam variant generation: A human-in-the-loop framework for automatic item creation. *Education Sciences*, *15*(8). https://doi.org/10.3390/educsci15081029

[4] Cappelleri, J. C., Lundy, J. J., & Hays, R. D. (2014). Overview of classical test theory and item response theory for quantitative assessment of items in developing patient-reported outcome measures. *PubMed Central*, *36*(5), *648–662*. https://doi.org/10.1016/j.clinthera.2014.04.006

[5] Cooperman, A. (2022). *An automated test assembly approach using item response theory to enhance evidence of measurement invariance [Ph.D. Dissertation, University of Minnesota].* https://hdl.handle.net/11299/241639

[6] Daher, W., Diab, H., & Rayan, A. (2023). Artificial intelligence generative tools and conceptual knowledge in problem solving in chemistry. *Information*, *14*(7), *409.* https://doi.org/10.3390/info14070409

[7] Dan, N., Cai, Y., & Wang, Y. (2025). Symbolic or Numerical? Understanding physics problem solving in reasoning LLMs. *arXiv*. https://arxiv.org/html/2507.01334v2

[8] Ferrero, F. (2024, July 1). *Item Response Theory Practice with R: A Tutorial. RPubs.* https://rpubs.com/fferrero/IRT_Tutorial

[9] Jindal, N. (2025). Upholding examination integrity: The Public Examinations (Prevention of Unfair Means) Act, 2024. *Journal of Emerging Technologies and Innovative Research (JETIR)*, *12*(3*), 643-645.* https://www.jetir.org/papers/JETIR2503981.pdf

[10] Laurent, A. (2026, February 15). *Wolfram Alpha vs ChatGPT: Comprehensive comparison of symbolic and generative AI. Intuition Labs.* https://intuitionlabs.ai/articles/symbolic-ai-vs-generative-ai-wolfram-chatgpt

[11] Lavin, R., Liu, X., Mohanty, H., Norman, L., Zaarour, G., & Krishnamachari, B. (2024). A survey on the applications of zero-knowledge proofs. *arXiv*. https://arxiv.org/html/2408.00243v1

[12] Michaelides, M. P. (2010). A review of the effects on IRT item parameter estimates with a focus on misbehaving common items in test equating. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2010.00167

[13] Oikonomou, O., Lingsch, L., Grund, D., Mishra, S., & Kissas, G. (2026). Neuro-symbolic AI for analytical solutions of differential equations. *arXiv*. https://arxiv.org/html/2502.01476v3

[14] Pommerich, M. (2016). *The fairness of comparing test scores across different tests or modes of administration (pp. 111-134). Routledge.* https://doi.org/10.4324/9781315774527-9

[15] Proietti, G. S., Matteucci, M., & Mignani, S. (2020). Automated test assembly for large-scale standardised assessments: Practical issues and possible solutions. *Psych*, *2*(4), *315-337.* https://doi.org/10.3390/psych2040024

[16] Ramakrishna, D., & Shaik, M. A. (2025). A comprehensive analysis of cryptographic algorithms: Evaluating security, efficiency, and future challenges. *IEEE Access*, *13*. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10804125

[17] Reddy, M. V., Sri, P. N., Vivek, R., Reddy, T. S. N., & Khan, H. Q. (2026). Blockchain-enhanced secure examination management system using MFA. *International Research Journal of Engineering and Technology (IRJET)*, *13*(4). https://www.irjet.net/archives/V13/i4/IRJET-V13I04154.pdf

[18] Shaikh, M. F., Hassan, S. H., Maccaro, A., Pratesi, G., & Piaggio, D. (2023). A blockchain framework using proof of authority and smart contracts for ethical and secure healthcare asset management. *Front Public Health*. https://doi.org/10.3389/fpubh.2025.1638546

[19] Sharma, S. (2025, December 17). From digital to paper: Can pen-and-paper exams end NTA's test turmoil and shape the future of JEE, NEET and CUET? *The Times of India*.

[20] Solomon, R. G., Sowmya, K. N., & Chennamma, H. R. (2024, December 21). *Security of examination question paper through blockchain - SecureQ [Conference presentation]. 2024 International Conference on Innovation and Novelty in Engineering and Technology (INNOVA).* https://ieeexplore.ieee.org/document/10847055

[21] Tan, L., Waters, C., Huang, F., & Cloney, D. (2024). Unpacking automated test assembly: New findings and directions for the future. *OECD Programme for International Student Assessment (PISA)*. https://doi.org/10.37517/2024041114-01

[22] Tejashwini, Y., Farzanula, M., Nikhil, S., & Khan, S. A. (2025). Question paper leakage protection using blockchain. *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering*, *13*(9). https://doi.org/10.17148/IJIREEICE.2025.13904

[23] Verghese, P. C. (2001, February 5). *Estimation of time to complete a print job. Google Patents.* https://patents.google.com/patent/EP1096364A2/en

[24] Want, R. (2011). Near field communication. *IEEE Pervasive Computing*, *10*(3), *4-7.* https://doi.org/10.1109/MPRV.2011.55

[25] Xu, Z., Ong, C. S., Kandanaarachchi, S., & Ntoutsi, E. (2024). Fairness evaluation with item response theory. *arXiv*. https://arxiv.org/pdf/2411.02414

[26] Yadav, H. K., Adithya, G., Ehshanulla, K., Srinivas, T. M., Yadav, M., & Gowrishankar, Y. (2025). Blockchain-powered question paper leakage prevention system. *International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025)*. https://doi.org/10.5220/0013888700004919