The Future of Income Assessment in Digital Lending

Kunal Urdhwareshe
Head – Analytics, Aditya Birla Housing Finance Limited, Mumbai, Maharashtra, India

***Abstract*—Income is the primary determinant of repayment capacity, yet India assesses it through documentary artefacts — salary slips, Form 16, income tax returns and manually reviewed bank statements — designed for a formal economy that a majority of Indian earners do not inhabit. This paper argues that income assessment is undergoing a structural transition from point-in-time verification, through analytical assessment, to continuous income intelligence: a probabilistic, explainable and continuously refreshed view of a borrower's sustainable debt-service capacity. It examines the limitations of documentary underwriting; proposes a five-dimensional construct of income comprising level, stability, predictability, resilience and trajectory, consolidated into an Income Intelligence Score reported with explicit confidence; and assesses the enabling technologies of machine learning, explainable artificial intelligence and generative artificial intelligence. It evaluates India's Account Aggregator framework, which has crossed 45 crore cumulative consents and 500 crore data fetches and received a formal self-regulatory anchor in 2026, alongside the Open Credit Enablement Network as the protocol layer for cash-flow-based credit. The paper analyses emerging data sources and their limitations, the regulatory perimeter established by the Reserve Bank of India (Digital Lending) Directions, 2025 and the Digital Personal Data Protection Rules, 2025, and the fairness risks inherent in income modelling. Ten case studies from India and comparable markets, three scenarios for 2030, and stakeholder-specific recommendations are presented. The central conclusion is that the binding constraint on inclusive credit growth in India is no longer algorithmic capability or data availability, but institutional willingness to rewrite credit policy around probabilistic income and to build the governance that makes such policy defensible.**

***Index Terms*—Account Aggregator, alternative data, artificial intelligence, cash-flow lending, credit risk, digital lending, financial inclusion, income intelligence, MSME finance, open finance, responsible AI.**

EXECUTIVE SUMMARY

Key findings

India has built, at population scale, the consent infrastructure required to observe income directly rather than infer it from documents. The Account Aggregator ecosystem has delivered over 45 crore cumulative consents and more than 500 crore data fetches, processes over 7 lakh consents daily, and spans more than 1,100 live regulated entities [6], [7]. Despite this, credit growth remains concentrated among borrowers already visible to the system: MSME commercial credit expanded 13% to ₹35.2 lakh crore in FY25 with delinquencies at a five-year low of 1.8%, yet the new-to-credit share of originations fell from 51% to 47% [3].

The gap is therefore not one of risk appetite or of data availability, but of measurement. Documentary income assessment cannot observe the self-employed, the micro-enterprise, the platform worker or the informally employed — populations that include a gig workforce projected to reach 23.5 million by 2029–30 [10] and an MSME sector facing an addressable credit gap estimated at ₹20–25 trillion [4].

Strategic implications

Institutions that treat consented data merely as a substitute for paper will capture process efficiency and little else. The larger value lies in redesigning credit policy around a probabilistic income estimate with an explicit confidence band, in extending consent pipelines from origination into continuous early-warning monitoring, and in distinguishing rigorously between borrowers assessed as low-capacity and borrowers who are simply poorly observed. The regulatory perimeter — the Digital Lending Directions, 2025 and the phased Digital Personal Data Protection regime culminating on 13 May 2027 — makes model governance, consent stewardship and explainability prerequisites for scale rather than optional overlays [1], [11].

Future outlook

By 2030, the base case is mainstream adoption of AI-based income estimation and open finance across leading banks, NBFCs and fintechs, with the largest institutions operating portable, continuously refreshed income intelligence in selected segments. The decisive differentiator will be labelled portfolio outcome data and validated models — not algorithms, which are commoditizing rapidly. The principal risk to manage is the Kenyan failure mode: data-enabled speed without affordability discipline, which produces rapid inclusion followed by rapid over-indebtedness.

I. INTRODUCTION

A. The Evolution of Income Assessment in Lending

Every credit decision, regardless of product, geography or vintage, ultimately rests on a single question: can this borrower pay, and will this borrower pay? Collateral, guarantees and covenants are secondary defences. Income — its level, its regularity, its durability and its trajectory — is the primary source of repayment. The history of lending is therefore, in large part, a history of successive attempts to observe income more cheaply, more accurately and more frequently.

India's journey through that history has been unusually compressed. Within a single generation, the market has moved from relationship-based village credit, to branch-based documentary underwriting built on salary slips and audited financials, to bureau-assisted scorecards after the establishment of credit information companies, and now to a digital-native architecture in which consented data flows, transaction streams and machine learning models are beginning to substitute for paper. The Reserve Bank of India (RBI) consolidated the rules for this last phase through the Reserve Bank of India (Digital Lending) Directions, 2025, which brought customer protection, data governance, technology standards and loss-sharing arrangements into a single instrument [1].

What has not changed is the centrality of income. What is changing — decisively — is how income is known. The thesis of this paper is that Indian lending is migrating from income verification (a point-in-time documentary check) through income assessment (an analytical estimate) towards income intelligence (a continuous, probabilistic, explainable view of a household's or enterprise's capacity to service debt). This migration is not merely a technology upgrade. It redefines the addressable market, the unit economics of origination, the shape of risk management and the boundaries of regulatory oversight.

B. Why Income Remains the Foundation of Underwriting

Three structural reasons keep income at the centre of credit. First, income determines affordability, and affordability determines loss given stress far more reliably than collateral in unsecured and semi-secured portfolios. Second, regulation increasingly codifies income-based limits: the RBI's microfinance framework, for example, requires an assessment of household income and caps monthly repayment obligations as a proportion of household income [2]. Third, income is the variable that most directly links credit growth to real economic activity; portfolios that grow faster than borrower incomes create systemic fragility irrespective of how strong bureau scores appear at origination.

The corollary is uncomfortable. If income is the foundation, and if income is observed poorly for a large share of the population, then a large share of credit decisions in India rest on an estimate whose error bars are neither measured nor disclosed.

C. The Indian Challenge

India's lending ecosystem presents an unusual combination: world-class digital public infrastructure coexisting with a predominantly informal labour market. The Micro, Small and Medium Enterprises (MSME) sector illustrates the tension. Commercial credit exposure to MSMEs reached approximately ₹35.2 lakh crore as of March 2025, growing 13% year-on-year, with balance-level serious delinquencies falling to a five-year low of 1.8% [3]. Growth was nonetheless driven largely by deeper lending to existing borrowers rather than by new borrower acquisition, and the share of new-to-credit MSMEs in new originations fell from 51% to 47% over the year [3]. The addressable credit gap for MSMEs is commonly estimated at ₹20–25 trillion, with formal credit penetration in the low double digits [4].

The pattern is familiar: the system is efficient at lending more to those it already understands, and structurally weak at understanding someone new. That weakness is, at its core, an income-observation problem.

II. THE LIMITATIONS OF TRADITIONAL INCOME ASSESSMENT

Traditional income assessment in India rests on four artefacts — the salary slip, the bank statement, the income tax return (ITR) and Form 16 — supplemented by field verification. Each was designed for a formal, salaried, documented economy. Each fails in predictable ways at the frontier.

A. Salary Slips and Form 16

These instruments are precise where they apply and silent where they do not. They cover only the formally employed with structured payroll, excluding the self-employed, the informally employed and the platform worker by construction. They are also static: a salary slip certifies a payment made, not an income likely to persist. A borrower in a sector facing contraction and a borrower in a sector expanding rapidly produce identical documents.

B. Bank Statements

Manually reviewed PDF statements remain the workhorse of self-employed underwriting. They are rich in signal but expensive to process, inconsistently formatted across institutions, easily manipulated in circulated-PDF form, and typically sampled over a short window — six to twelve months — that cannot separate seasonality from decline. Where a borrower banks with multiple institutions, single-statement analysis systematically misstates both inflows and obligations.

C. Income Tax Returns

ITRs are authoritative for the tax base but a weak proxy for cash flow. For the small business owner, reported taxable income is shaped by presumptive taxation, depreciation and legitimate tax planning; the number that matters for repayment — free cash flow available to the household — can be a multiple of the number that appears on the return. ITRs are also stale by design, arriving with a lag of months to over a year.

D. Manual Verification and Fraud

Physical income verification is slow, costly, non-reproducible and susceptible to collusion. Document-based fraud — fabricated slips, edited statements, synthetic employers — is among the most persistent fraud typologies in retail and MSME lending precisely because the artefact, not the underlying cash flow, is the object of verification. Fetching data directly from the source institution under consent substantially reduces this attack surface.

E. Exclusion and Scalability

The deepest limitation is distributional. A documentary regime prices out precisely those segments where credit would be most transformative: micro-enterprises, gig and platform workers, agricultural and allied households, first-generation entrepreneurs and women borrowers with limited documentation in their own name. It is also economically unscalable: the cost of assessing a ₹30,000 working-capital loan through a documentary process exceeds any plausible margin on that loan. Small-ticket, short-tenure credit is not unviable; it is unviable under documentary underwriting.

TABLE I. TRADITIONAL VERSUS EMERGING INCOME ASSESSMENT

Dimension	Traditional Approach	Income Intelligence Approach
Primary input	Salary slip, ITR, Form 16, PDF bank statement	Consented transaction streams, GST, payroll and platform APIs
Nature of view	Point-in-time snapshot	Continuous, longitudinal
Output	Binary verified or not verified	Probabilistic income estimate with confidence interval
Turnaround	Days to weeks	Minutes to hours
Fraud exposure	High (document manipulation)	Low (source-fetched, digitally signed)
Marginal cost per case	High, largely manual	Low, largely automated
Coverage	Formal, salaried, documented	Formal and informal, including thin-file
Risk monitoring	Event-driven, post-delinquency	Continuous, pre-delinquency
Explainability	Human judgement, poorly recorded	Model-based, auditable, reason-coded

Dimension	Traditional Approach	Income Intelligence Approach
Regulatory posture	Document retention	Consent, purpose limitation, model governance

III. THE RISE OF INCOME INTELLIGENCE

Income intelligence is the discipline of inferring, from consented behavioral and transactional data, not only how much a borrower earns but how dependable, durable and directionally stable that earning is — and of maintaining that inference over the life of the exposure.

A. From Verification to Intelligence

Verification asks: is the stated income true? Assessment asks: what is the income? Intelligence asks a richer question: what is the distribution of this borrower's future disposable cash flow, how confident are we, and what would move it? The difference is the difference between a photograph and a moving image with a forecast attached.

B. The Five Dimensions

Practical implementation requires decomposing income into measurable dimensions. The author proposes five that are both estimable from Account Aggregator (AA)-grade data and economically meaningful.

- Income level: the central estimate of recurring monthly inflow attributable to economic activity, net of circular transfers, loan disbursements and self-transfers.
- Income stability: dispersion of that inflow across time, typically a coefficient of variation computed over rolling windows, adjusted for seasonality.
- Income predictability: the extent to which the next period's inflow can be forecast from history — measured by out-of-sample forecast error rather than by variance alone. A seasonal business can be highly variable yet highly predictable.
- Income resilience: the ability to absorb a shock, proxied by liquidity buffers, expense compressibility, income-source diversification and the behavior of inflows during past disruptions.
- Income trajectory: the direction and momentum of income over twelve to twenty-four months, which distinguishes a borrower at ₹40,000 and rising from a borrower at ₹40,000 and falling — two profiles identical to a salary slip and materially different in risk.

C. Cash-Flow Behavior Analytics

Surrounding these dimensions is a wider behavioral layer: the ratio of fixed obligations to inflow; end-of-cycle liquidity; frequency of bounced mandates; concentration of counterparties; the balance between digital and cash settlement; and the timing relationship between inflows and obligations. Behavioral indicators frequently carry more predictive power than the income estimate itself, because they reveal how a household actually manages the money it has.

D. The Income Intelligence Score

The author proposes consolidating these into an Income Intelligence Score (IIS): a composite, lender-agnostic measure expressing sustainable debt-service capacity together with the confidence attached to that estimate. Conceptually,

IIS = f (level, stability, predictability, resilience, trajectory; confidence), reported as a score paired with an explicit confidence band and a set of reason codes.

Two design principles matter. First, the score must report its own uncertainty: a ₹50,000 estimate drawn from twenty-four months of full-coverage data is a different underwriting object from the same estimate drawn from four months of partial data, and collapsing both into one number destroys information the credit committee needs. Second, the score must be decomposable. A single opaque number invites the same regulatory objection as any black-box model; a score that resolves into five named, auditable dimensions supports both supervisory review and borrower-facing explanation.

Infographic recommendation 1 — "The Income Intelligence Radar": a five-axis radar chart plotting level, stability, predictability, resilience and trajectory for three archetypes (salaried professional, kirana owner, delivery partner), demonstrating that identical income levels produce materially different risk shapes.

IV. AI-POWERED INCOME ASSESSMENT

Artificial intelligence enters income assessment not as a replacement for credit judgement but as an instrument for extracting structure from data volumes that human underwriting cannot process.

A. Transaction Classification and Income Estimation

The foundational task is classification: assigning each transaction in a consented statement to a category — salary, business receipt, interest, rental, transfer, refund, loan disbursement, gambling, and so on. This is a supervised learning problem over sparse, noisy, abbreviated narration strings, where modern text-embedding and gradient-boosting ensembles substantially outperform rule-based keyword matching, particularly on regional-language and merchant-abbreviated narrations. Accuracy at this layer governs everything downstream: a misclassified inter-account transfer inflates estimated income and silently mis-prices the loan.

Above classification sits the estimation model, which maps classified flows to a recurring-income estimate. The output should be a distribution, not a scalar. Regression models with quantile outputs, or ensembles producing predictive intervals, allow policy to be written against the lower bound for conservative products and the central estimate for prime segments.

B. Pattern Recognition and Behavioral Analytics

Sequence models identify salary-day regularity, seasonality in trade receipts, festival-linked spikes and creeping deterioration. Anomaly detection flags synthetic patterns characteristic of statement manipulation — round-figure inflows, implausible balance arithmetic, or inflow clusters immediately preceding a loan application. Graph analytics over counterparties surfaces circular-transaction rings designed to manufacture turnover.

C. Affordability Assessment

Income intelligence permits affordability to be computed rather than assumed. Instead of applying a uniform fixed-obligation-to-income ratio, the lender can estimate a borrower-specific surplus: estimated recurring income, less observed recurring obligations, less an estimate of non-compressible household expenditure, stressed for income volatility. This produces a defensible answer to the supervisory question of whether the borrower can service the loan without distress — a question that fixed ratio policies answer only on average.

D. Explainable AI

Explainability is not a compliance ornament in credit; it is an operating requirement. Adverse-action reasoning, internal validation, and the RBI's broader expectations on model governance and accountability under the FREE-AI framework all require that a decision be reconstruct able [5]. Practical approaches include monotonic constraints on economically signed features, Shapley-value attributions at case level, surrogate scorecards for policy articulation, and reason-code libraries mapped to borrower-comprehensible language. Where a model cannot be explained to a branch credit manager, it will not survive a supervisory review.

E. Generative AI

Generative models add three distinct capabilities. They normalize unstructured artefacts — scanned financials, GST annexures, rent agreements, informal ledgers — into structured fields. They draft underwriting narratives and deviation notes from structured model output, compressing credit-memo preparation time. And they support conversational data collection, allowing a borrower to describe income in natural language in a regional language, with the model mapping the description to verifiable data requests. The governance caution is equally clear: generative output should inform the file, not determine the decision, and hallucination risk makes unconstrained generative underwriting inappropriate.

F. Real-Time Risk Monitoring

The most under-exploited application is post-disbursement. Where consent permits periodic refresh, income intelligence converts underwriting from an event into a process: salary-credit cessation, sustained inflow decline, rising bounce frequency or the appearance of new obligations become early-warning triggers weeks before a missed instalment. This enables pre-emptive restructuring, limit modulation and prioritized collections — a shift from reactive recovery to anticipatory portfolio management.

TABLE II. LENDING USE CASES FOR AI-DRIVEN INCOME INTELLIGENCE

Use case	Mechanism	Business outcome
Instant personal loan decisioning	AA fetch plus transaction classification and income model	Lower turnaround, reduced drop-off

Use case	Mechanism	Business outcome
Self-employed home loan assessment	Multi-account cash-flow reconstruction and surplus modelling	Expanded approvals with controlled risk
MSME working capital	GST turnover triangulated with bank inflows	Cash-flow-based limits replacing collateral tests
Gig worker credit	Platform payout streams and earnings volatility modelling	Access for a previously unbankable segment
Portfolio early warning	Periodic consent refresh with inflow monitoring	Earlier intervention, lower credit cost
Limit management	Continuous affordability recomputation	Dynamic, income-linked exposure control
Fraud control	Anomaly and counterparty graph analytics	Reduction in documentary fraud losses
Collections prioritisation	Capacity-to-pay segmentation	Higher recovery per rupee of effort

V. OPEN FINANCE AND THE ACCOUNT AGGREGATOR REVOLUTION

India's Account Aggregator framework is the single most consequential development for income assessment in the country's lending history. It converts financial data from an artefact the borrower must produce into a right the borrower can exercise.

A. Architecture

The AA framework is a consent-mediated data-sharing network regulated by the RBI. Financial Information Providers (FIPs) — banks, non-banking financial companies, insurers, depositories, GST systems and pension repositories — hold data. Financial Information Users (FIUs) — lenders, wealth managers, insurers — consume it. Between them sit licensed Account Aggregators, which are data-blind consent intermediaries: they route consent artefacts and encrypted payloads without reading the data. Consent is granular, specifying purpose, data range, frequency and duration, and is revocable by the customer at any time.

B. Scale Achieved

The framework has crossed the threshold from pilot to production infrastructure. Ecosystem statistics published by Sahamati indicate more than 45 crore cumulative consents fulfilled and over 500 crore cumulative data fetches, with the network processing over 7 lakh consents daily and facilitating close to 3.8 crore financial products and services in FY26 [6]. The network comprises over 1,100 live regulated entities, including 176 FIPs, 1,020 FIUs and 17 operational Account Aggregators, against roughly 294 million linked accounts [7]. In June 2026 the RBI recognized Sahamati Foundation as the Self-Regulatory Organization for the AA ecosystem, giving a formerly voluntary industry protocol a formal regulatory anchor [7], [8].

Adoption has also broadened beyond lending. In capital markets, an estimated 70–80% of income verification at leading brokers is now conducted through the AA network, and roughly 67.65 lakh futures and options accounts were verified via AA in FY26 [6]. This matters for lenders: it demonstrates that consent-based income verification can operate at national scale, at low marginal cost, in a regulated segment where the verification must withstand audit.

C. Data Portability and Real-Time Underwriting

Three lending consequences follow. First, portability breaks the informational monopoly of the borrower's primary banker, allowing any regulated lender to compete on the same evidence — which compresses pricing power built on information asymmetry rather than on cost of funds or service. Second, recurring consent enables monitoring, not merely origination; the same pipe that underwrites can supervise. Third, latency collapses: a data pull that once took a fortnight of document collection now takes seconds, which is what makes small-ticket lending economically viable.

D. Constraints

Candour is warranted. FIP coverage and technical uptime remain uneven; consent success rates vary materially by aggregator and by FIP; customer comprehension of what is being consented to is imperfect; and the framework observes only the formal financial system — an income received and held in cash remains invisible. The framework is necessary but not sufficient for full inclusion.

Chart recommendation 2 — a dual-axis line chart plotting monthly cumulative consents fulfilled and cumulative linked accounts from FY23 to FY26, annotated with the SRO recognition milestone, illustrating the inflection from pilot to infrastructure.

VI. DIGITAL LENDING FOR THE INVISIBLE BORROWER

The economic prize of income intelligence is not efficiency in serving customers already served. It is the extension of formal credit to borrowers who are creditworthy but unobservable.

A. Self-Employed Households

The self-employed constitute the largest under-served cohort in secured and unsecured retail. Their income is real, frequently substantial, and poorly documented. Cash-flow reconstruction across multiple accounts, combined with GST and platform receipts, allows lenders to build an evidenced income view where the ITR understates and the salary slip does not exist. For affordable housing finance specifically, the binding constraint on approval rates is rarely credit appetite; it is income evidence.

B. MSMEs

MSME credit shows the shape of the problem clearly. With commercial exposure at ₹35.2 lakh crore and serious delinquencies at a five-year low of 1.8%, portfolio quality is not the constraint [3]. Yet new-to-credit share in originations declined to 47% [3], and the addressable gap remains in the ₹20–25 trillion range with formal credit reaching a minority of enterprises [4]. The Open Credit Enablement Network (OCEN) was designed precisely for this gap: an interoperable protocol that unbundles lending into sourcing, underwriting, capital and collections roles, enabling small-ticket, short-tenure, cash-flow-linked credit delivered inside the platforms MSMEs already use [4], [9].

C. Gig and Platform Workers

NITI Aayog estimated 7.7 million gig workers in 2020–21, projected to reach 23.5 million by 2029–30, equivalent to 6.7% of the non-farm workforce, and explicitly recommended accelerating access to finance through products designed for platform workers [10]. This cohort is paradoxically the most data-rich and least credit-served population in India: every rupee earned is digitally recorded by the platform, yet the absence of an employer-issued salary slip renders it invisible to documentary underwriting. Platform earnings APIs and AA-fetched settlement credits together make gig income among the most verifiable income types in the economy.

D. Thin-File and New-to-Credit Consumers

For borrowers with no bureau history, income intelligence substitutes evidence of capacity for evidence of past repayment. The correct construct is not to force thin-file applicants through scorecards calibrated on thick-file behavior, but to build parallel decisioning where cash-flow features carry the weight that bureau features carry elsewhere — with correspondingly conservative initial limits and rapid, behavior-based limit escalation.

TABLE III. SEGMENT-WISE INCOME OBSERVABILITY AND ENABLERS

Segment	Traditional observability	Primary income-intelligence enabler
Salaried formal	High	Payroll APIs, AA salary credits
Self-employed professional	Medium	Multi-bank cash flow, GST, ITR triangulation
Micro-enterprise / kirana	Low	UPI and QR settlements, GST, POS data
Gig / platform worker	Very low	Platform payout APIs, settlement credits
Informal wage worker	Very low	Bank credits, utility and rent regularity
Agri and allied	Very low	Seasonal inflows, mandi receipts, subsidy credits
New-to-credit urban	Low	Cash-flow surplus with staged limit escalation

VII. EMERGING DATA SOURCES FOR INCOME ASSESSMENT

No single source is sufficient. The discipline lies in knowing what each source proves, what it cannot prove, and how sources triangulate.

A. Bank Transaction Data

The richest single source, and through AA the most reliably obtained. Strengths: granularity, obligation visibility, behavioral depth, tamper resistance when source-fetched. Limitations: multi-banking fragmentation, cash invisibility, and classification error on sparse narration.

B. GST Data

For registered enterprises, GST returns provide near-authoritative, periodically refreshed turnover with counterparty structure. Strengths: reliability, frequency, sectoral comparability. Limitations: turnover is not profit; coverage excludes unregistered micro-enterprises; and filing lags and nil-filing behavior require careful handling.

C. UPI and Digital Payment Behavior

UPI settlement flows are the closest available proxy to real-time revenue for micro-merchants. Strengths: high frequency, wide penetration, low fabrication risk when observed through the settlement account. Limitations: mixing of personal and business flows, refund and reversal noise, and the risk of double-counting when settlements also appear as bank credits.

D. E-Commerce and Marketplace Seller Data

Order and settlement histories from marketplaces provide verified revenue, return rates and ratings. Strengths: directly tied to a monitorable receivable and often to an escrow-controlled repayment path. Limitations: platform concentration risk and partial visibility where the seller operates across channels.

E. Payroll APIs

Direct payroll or employer-system integration provides authoritative gross and net pay with employment status. Strengths: precision and fraud resistance. Limitations: limited penetration outside organized employment and dependence on employer participation.

F. Accounting Software Data

Ledger-level data from accounting platforms provides receivables ageing, payables and margin structure. Strengths: analytical depth for small enterprises. Limitations: self-maintained and therefore manipulable; quality varies widely.

G. Telecom and Utility Signals

Recharge regularity, postpaid tenure, electricity consumption and rent payment history carry weak but non-trivial signal on stability, particularly for thin-file borrowers. These should be treated as supporting evidence and used with particular care on fairness grounds, since such variables can proxy for protected characteristics.

H. Embedded Finance Ecosystems

Where credit is originated inside a commercial platform — a distributor network, a fleet aggregator, a billing application — the platform observes both income generation and, frequently, controls the repayment path. This combination of observation and control is the most powerful configuration in digital lending, and also the one that most requires governance discipline on conflicts of interest and on the conduct requirements applicable to lending service providers under the 2025 Directions [1].

TABLE IV. DATA SOURCE ASSESSMENT MATRIX

Source	Signal strength	Coverage	Key limitation
Bank transactions (AA)	High	Broad	Cash invisibility; multi-banking
GST returns	High	Registered only	Turnover ≠ income
UPI / QR settlements	Medium-high	Very broad	Personal-business mixing
Marketplace seller data	High	Platform-bound	Concentration risk
Payroll APIs	Very high	Organized sector	Employer participation
Accounting software	Medium	Small	Self-reported

Source	Signal strength	Coverage	Key limitation
Telecom / utility	Low-medium	Very broad	Proxy and fairness risk
Platform payout APIs	High	Gig cohort	Platform dependence

Infographic recommendation 3 — "The Income Evidence Stack": a layered pyramid with authoritative sources (payroll, GST) at the apex, transactional sources (bank, UPI, marketplace) in the centre, and inferential signals (telecom, utility) at the base, with confidence weighting shown on the right edge.

VIII. RESPONSIBLE AI, PRIVACY AND REGULATORY CONSIDERATIONS

Income intelligence concentrates two sensitivities simultaneously: it processes intimate financial behavior, and it uses that processing to allocate access to credit. Both attract regulatory attention, and rightly so.

A. The DPDP Framework

The Digital Personal Data Protection Act, 2023 was operationalized through the DPDP Rules, 2025, notified on 13 November 2025 after a consultation that received 6,915 inputs [11]. The framework commences in phases: the Data Protection Board was constituted in November 2025; Consent Manager registration provisions take effect from November 2026; and the balance of substantive obligations — notice, consent, security safeguards, breach notification, data principal rights and Significant Data Fiduciary duties — take effect from 13 May 2027, with penalties of up to ₹250 crore for security failures leading to breach [11], [12].

For lenders, four operational implications dominate. Purpose limitation means income data fetched for underwriting cannot be silently repurposed for cross-sell. Data minimisation means the twenty-four-month full-statement pull convenient for modelling must be justified against the decision being made. Retention discipline means erasure schedules must be engineered into the data platform, not promised in policy. And the architectural convergence of the AA consent artefact with the DPDP Consent Manager construct suggests that consent, not contract, will be the durable basis for financial data processing in India.

B. RBI Expectations

The Digital Lending Directions, 2025 require regulated entities to assess borrower creditworthiness in an auditable manner, to conduct enhanced due diligence on lending service providers, to obtain explicit consent before sharing personal information with third parties, to refrain from storing biometric data, to maintain comprehensive privacy policies, to report digital lending applications to the RBI through the CIMS portal, and to cap default loss guarantee cover at 5% of the disbursed portfolio [1], [13]. The direction of travel is clear: the regulator accepts technological delegation of process but not delegation of accountability.

Beyond digital lending specifically, the RBI's framework for responsible and ethical enablement of artificial intelligence articulates principles — trust as the foundation, people-first design, innovation with prudence, fairness, accountability and explainability — that boards should treat as the reference standard for income models [5].

C. Algorithmic Fairness and Bias

Income models can encode disadvantage in three ways. Historical bias: training on approved-population outcomes reproduces the exclusions of the incumbent process. Proxy bias: geography, device, language or merchant category can stand in for caste, religion, gender or region without any protected attribute appearing in the model. Measurement bias: income is observed with systematically greater error for informal earners, and if model confidence is not modelled explicitly, that error is silently converted into rejection.

Mitigation is operational, not rhetorical: reject-inference and controlled random-approval strategies to correct selection bias; disparate-impact testing across gender, geography and occupation cohorts at both approval and pricing stages; proxy-variable audits; and — critically — separating "we assess this borrower as low capacity" from "we cannot observe this borrower well", because the two demand different treatments. The first is a credit decision; the second is a data-coverage problem that should trigger alternative evidence collection rather than rejection.

D. Transparency, Governance and Consent

A defensible governance architecture for income models includes: model tiering by materiality;

independent validation separated from development; documented data lineage from consent artefact to feature; challenger models and champion-challenger monitoring; drift detection on both inputs and outcomes; human override with recorded reasoning; and board-level reporting on model performance and fairness outcomes. Consent design deserves equal attention: comprehension, not merely click-through, is the standard the framework aspires to, and lenders that treat consent as a legal formality will find it the weakest link in their compliance chain.

IX. CASE STUDIES

The following cases are drawn from public information and are presented as illustrations of mechanism rather than as endorsements.

1) GeM Sahay (India). The first OCEN pilot enabled purchase-order financing for sellers on the Government e-Marketplace, with fully digital journeys from remote lenders using an auction model. Public reporting indicates disbursement of over ₹23 crore to more than 14,000 registered sellers as of January 2024, with the smallest transaction at ₹160 [9], [14]. The lesson is dual: cash-flow-linked micro-credit is technically feasible at extreme small-ticket economics, and the repayment rail matters as much as the underwriting — lenders subsequently sought stronger platform support for collections [14].

2) GST Sahay (India). Invoice-based financing for MSMEs using GST filings as the income substrate demonstrated that a government-mandated compliance artefact can be repurposed as credit evidence, shortening the distance between a sale being made and working capital being available [9].

3) Account Aggregator income verification in capital markets (India). Brokers now conduct an estimated 70–80% of income verification through the AA network, with approximately 67.65 lakh derivatives accounts verified in FY26 [6]. This is the clearest existing proof that consent-based income verification can replace documentary verification at national scale in a regulated, audited process.

4) UPI-linked merchant lending (India). Lenders and payment platforms underwrite micro-merchants on observed QR settlement volumes, with repayment collected as a share of daily settlements. The case illustrates the power of combining income observation with repayment control, and the concentration risk that arises when both sit with one platform.

5) Gig platform earnings-based credit (India). Ride-hailing and delivery platforms have partnered with NBFCs to offer vehicle, device and personal credit underwritten on platform earnings history. Against NITI Aayog's projection of 23.5 million gig workers by 2029–30 [10], this is a template for converting the most digitally observable workforce in India into a bankable one.

6) Cash-flow-based MSME lending on GST plus bank triangulation (India). Several NBFCs now price unsecured business loans on GST turnover reconciled against AA-fetched bank credits, using divergence between the two as a fraud and quality signal rather than as a disqualifier. The mechanism converts two individually imperfect sources into one reasonably robust estimate.

7) Open Banking in the United Kingdom. The UK's regulated open banking regime established that mandated API access plus a certified participant framework produces reliable affordability data at scale, and that adoption follows utility rather than mandate alone. India's AA architecture, being consent-led and multi-sector from inception, is arguably broader in scope than the UK's payment-account-centred origin.

8) Cash-flow underwriting in United States consumer lending. US lenders and data networks have demonstrated that bank-transaction-derived cash-flow variables add predictive power incremental to bureau scores, particularly for thin-file and near-prime applicants — an empirical result directly transferable to India's new-to-credit cohort.

9) Kenya's mobile-money-based lending. Mobile-money transaction histories were used to underwrite micro-loans at population scale, proving that alternative data can create credit access rapidly — and, through subsequent over-indebtedness and multiple-borrowing episodes, proving equally that data-enabled speed without affordability discipline creates

consumer harm. This is the cautionary case India should study most closely.

10) Brazil's open finance rollout. A phased, regulator-driven expansion from account data to transactional and investment data illustrates the sequencing benefit of mandating FIP participation early — a contrast with India, where uneven FIP readiness has been a practical constraint on consent success rates.

X. FUTURE SCENARIOS: INCOME ASSESSMENT IN 2030

A. Scenario One — Conservative Evolution

AA adoption continues but plateaus as a document-substitution tool. Lenders use consented statements to replace PDFs without re-architecting credit policy; income remains a point-in-time input to unchanged scorecards. Model governance is compliance-driven. New-to-credit share stays broadly flat, inclusion gains are incremental, and competitive differentiation reverts to distribution and cost of funds. In this world, the technology is adopted and the business model is not; the principal beneficiaries are process-efficiency lines, not growth lines.

B. Scenario Two — Transformational Adoption

AI-based income estimation and open finance become mainstream across the top tier of banks, NBFCs and fintechs. Cash-flow-based limits replace collateral-anchored limits in significant parts of MSME and self-employed lending; OCEN-style embedded credit scales inside commercial platforms; and periodic consent refresh makes portfolio monitoring continuous for a meaningful share of exposures. Approval rates for thin-file segments rise materially without proportionate credit-cost deterioration, because affordability is computed rather than assumed. Regulatory attention shifts from whether models may be used to how they are governed. Competitive advantage concentrates among institutions with proprietary data pipelines and validated models.

C. Scenario Three — The Intelligent Credit Ecosystem

Continuous income intelligence becomes shared underwriting infrastructure. A portable, consent-governed income representation — an Income Intelligence Score with published methodology and confidence bands — travels with the borrower across lenders, analogous to the bureau score but capacity-based rather than history-based. Credit becomes a persistent, dynamically sized facility rather than a discrete transaction: limits adjust as income evolves, pricing reflects observed rather than assumed volatility, and distress is anticipated rather than discovered. Underwriting as a distinct organizational function contracts; model risk management, data governance and consent stewardship expand.

The implications differ sharply by constituency. Banks gain from data depth and funding cost but must overcome legacy decisioning stacks. NBFCs gain the most in relative terms, since income intelligence neutralizes the physical-distribution advantage of incumbents. Fintechs shift from arbitrage on speed to competition on model quality. Regulators must supervise models and data flows rather than documents and branches — a capability shift as significant as the industry's. Consumers gain access and, if governance holds, fairer pricing; if it does not, they face opaque decisions and the over-indebtedness risk that speed without affordability discipline reliably produces.

The author's assessment is that Scenario Two is the base case for 2030, with the largest ten to fifteen lenders operating at the boundary of Scenario Three in selected segments. The binding constraint will not be algorithmic capability; it will be institutional willingness to rewrite credit policy around probabilistic income and to build the governance that makes such policy defensible.

Chart recommendation 4 — a three-column scenario comparison showing, for 2030, indicative ranges for new-to-credit origination share, average decision turnaround, share of portfolio under continuous monitoring and model-governance intensity under each scenario.

XI. STRATEGIC RECOMMENDATIONS

A. For Banks

Short term (0–12 months): make AA the default income-evidence channel across retail and MSME journeys, with documentary collection as the exception rather than the norm; instrument consent success rates by FIP and aggregator, and manage them as a funnel metric; establish an income-model inventory under formal model risk management.

Medium term (1–3 years): rebuild affordability policy around computed borrower-level surplus rather than uniform obligation ratios; deploy periodic consent refresh for early-warning monitoring on at least the unsecured and self-employed books. Long term (3–5 years): treat income intelligence as an enterprise data asset shared across credit, collections, cross-sell and provisioning, and integrate cash-flow features into expected credit loss staging.

B. For NBFCs

Short term: priorities segments where documentary underwriting is most punitive — self-employed affordable housing, micro-LAP, gig and platform credit — since that is where income intelligence produces the largest incremental approval rate per unit of incremental risk. Medium term: build proprietary classification and income-estimation models on own-portfolio outcome data rather than relying solely on vendor scores; the differentiating asset is labelled performance data, not the algorithm. Long term: convert episodic lending relationships into monitored, dynamically sized facilities, and use continuous income visibility to lower credit cost as a structural offset to higher funding cost.

C. For Fintechs

Short term: ensure that every lending service provider arrangement is squarely compliant with the 2025 Directions on disclosures, data handling and loss-sharing limits [1]. Medium term: specialize — in a protocol world, defensibility comes from being the best at one layer (classification accuracy, gig-income modelling, collections intelligence) rather than from replicating the full stack. Long term: build for interoperability with OCEN-style networks and consent-manager infrastructure rather than around proprietary data captivity, which regulation is steadily eroding.

D. For Policymakers

Short term: accelerate FIP coverage and enforce technical service-level standards, since uneven FIP availability is the principal practical brake on consent-based underwriting. Medium term: extend income-relevant data availability — payroll, platform earnings, utility — into the consent architecture under appropriate safeguards; align the DPDP Consent Manager framework with the AA consent artefact to avoid parallel, incompatible consent regimes. Long term: consider a public-interest standard for income-estimation disclosure, so that borrowers can see and contest the income a lender has inferred for them.

E. For Regulators

Short term: issue supervisory expectations specific to income-estimation models — validation, documentation, adverse-action explainability and fairness testing — building on the responsible AI principles already articulated [5]. Medium term: develop supervisory capacity to examine models and data lineage directly, including the ability to replay decisions. Long term: guard against the systemic risk that arises when the industry converges on a small number of similar income models and correlated data sources, which can synchronize credit expansion and contraction across lenders.

F. For Technology Providers

Short term: publish accuracy benchmarks for transaction classification and income estimation by segment, with confusion characteristics disclosed; unverifiable accuracy claims are the sector's principal credibility problem. Medium term: ship explainability, reason codes, drift monitoring and audit trails as core product, not as add-ons. Long term: build to open standards so that lenders can switch providers without re-architecting — which is also the fastest route to institutional adoption at scale.

TABLE V. PRIORITY ROADMAP BY HORIZON

Horizon	Core objective	Representative actions
0–12 months	Evidence substitution	AA-first journeys; model inventory; consent funnel analytics
1–3 years	Decision redesign	Computed affordability; continuous monitoring; proprietary models
3–5 years	Infrastructure shift	Portable income intelligence; dynamic limits; ECL integration

XII. KEY TAKEAWAYS FOR SENIOR LEADERS

1) Income, not collateral or bureau history, is the binding constraint on inclusive credit growth in India — and it is the variable the industry currently measures least well.
2) The infrastructure question is settled. With over 45 crore cumulative consents and 500 crore data fetches, and a recognized self-regulatory organization in place, Account Aggregator is production infrastructure, not an experiment [6], [7].
3) Document substitution is not transformation. Using consented data to replace PDFs while leaving credit policy unchanged captures perhaps a fifth of the available value.
4) Probabilistic income beats binary verification. An income estimate paired with a confidence band supports better pricing, better limits and a more defensible audit trail than a verified or not-verified flag.
5) The largest under-served pools are the most digitally observable: gig workers projected at 23.5 million by 2029–30 [10], and MSMEs facing a ₹20–25 trillion credit gap [4].
6) Data coverage failure is not the same as creditworthiness failure. Conflating the two converts a technology limitation into financial exclusion.
7) Governance is a growth enabler, not a tax. Under the 2025 Directions and the phased DPDP regime culminating in May 2027, institutions without model and consent governance will face capacity constraints on scaling, not merely compliance findings [1], [11].
8) Competitive advantage will accrue to labelled outcome data, not to algorithms. Models are increasingly commoditized; performance history on your own portfolio is not.
9) Continuous monitoring is the underexploited profit pool. Most institutions have built consent pipes for origination and have not extended them to early warning, where credit-cost savings are typically larger.
10) The failure mode to avoid is Kenya's: data-enabled speed without affordability discipline produces rapid inclusion followed by rapid over-indebtedness and regulatory retrenchment.

XIII. QUESTIONS EVERY LENDING INSTITUTION SHOULD ASK TODAY

1. What proportion of our originations use consented data as the primary income evidence, and what is our consent success rate by aggregator and by FIP?
2. Do we know the accuracy of our income estimates? Have we back-tested estimated income against realized repayment capacity by segment?
3. Does our credit policy consume a point estimate or a distribution? Where in policy does model confidence change the decision?
4. Can we distinguish, in our decline reasons, between insufficient capacity and insufficient data — and do the two-follow different operational paths?
5. How many applicants do we reject for documentation reasons who would be approved on cash-flow evidence? What is the annual revenue value of that population?
6. Is our income intelligence used after disbursement? What share of the portfolio is under periodic consent-based monitoring?
7. Who independently validates our income models, and is that function organizationally separate from model development?
8. Have we tested our models for disparate impact across gender, geography and occupation at both approval and pricing?
9. Can we explain any individual decision to the borrower, to our board and to a supervisor — and can we replay it on the data as it stood at decision time?
10. Are our consent artefacts, retention schedules and purpose limitations DPDP-ready ahead of the May 2027 enforcement date [11]?
11. If a competitor approved our declined applicants using better income intelligence, what would happen to our portfolio mix over three years?

XIV. CONCLUSION

The trajectory is now legible. Income verification asked whether a document was genuine. Income assessment asked what the number was. Income intelligence asks what a borrower can sustainably afford, how confident we are, and what is changing — continuously, and with reasons that can be stated.

India is unusually well positioned for this transition. It possesses consent infrastructure operating at population scale, a protocol layer designed for small-ticket cash-flow credit, a statutory privacy framework entering force, an articulated regulatory position on responsible AI, and a vast population of creditworthy but undocumented earners. Few markets have assembled these elements simultaneously.

What remains is institutional. The technology to estimate income from consented cash flow exists and works. The constraint is the willingness of credit committees to replace a familiar false precision — the salary slip, the ITR — with an honest probabilistic estimate, and to build the governance that makes such estimates defensible to a board and a supervisor. Institutions that make that shift will find that the population they can serve responsibly is considerably larger than their current book suggests. Those that do not will continue lending confidently to the customers they can already see, in a market where the growth is increasingly among those they cannot.

The next phase of Indian credit growth will not be won by the lender with the most branches or the cheapest funds. It will be won by the lender that understands income best.

XV. APPENDIX A. EXECUTIVE ONE-PAGE SUMMARY FOR THE BOARD

Issue

Income is the primary determinant of repayment capacity, yet it is assessed through documents designed for a formal economy that a majority of Indian earners do not inhabit. This constrains growth, concentrates risk in already-served segments and excludes creditworthy borrowers.

What has changed

Account Aggregator has moved from pilot to infrastructure — over 45 crore cumulative consents, over 500 crore data fetches, more than 1,100 live regulated entities, and RBI recognition of a self-regulatory organization in June 2026 [6], [7]. Machine learning can now convert consented transaction streams into income estimates with measurable accuracy. The regulatory perimeter is defined by the Digital Lending Directions, 2025 and the DPDP Rules, 2025 [1], [11].

The opportunity

MSME credit gap of ₹20–25 trillion with formal penetration in the low double digits [4]; new-to-credit MSME share in originations at 47% and declining [3]; gig workforce projected at 23.5 million by 2029–30 [10]. These segments are under-served because they are unobservable, not because they are uncreditworthy.

The proposition

Move from income verification to an Income Intelligence Score measuring level, stability, predictability, resilience and trajectory, reported with explicit confidence and reason codes, refreshed through the life of the exposure.

What the board should require

- An AA-first origination target with consent success rate reported as a board metric.
- Back-tested accuracy of income estimates by segment, presented at least annually.
- Credit policy that consumes model confidence, not only the point estimate.
- Independent model validation, fairness testing and full decision replay capability.
- DPDP readiness ahead of 13 May 2027, including consent, retention and purpose-limitation controls [11].
- Extension of consent pipelines from origination to continuous early-warning monitoring.

Principal risks

Model risk and unvalidated vendor accuracy claims; algorithmic bias and proxy discrimination; over-indebtedness from speed without affordability discipline; consent and privacy failures; and correlated-model systemic risk as the industry converges on similar data and methods.

ACKNOWLEDGMENT

The author thanks practitioners across India's banking, NBFC and fintech ecosystem whose operating experience informed the framing of this paper. The views expressed are personal and do not represent the position of any institution.

REFERENCES

[1] Reserve Bank of India. (2025). Reserve Bank of India (Digital Lending) Directions, 2025 (RBI/2025-26/36,

DOR.STR.REC.19/21.07.001/2025-26). Reserve Bank of India. https://www.rbi.org.in/scripts/NotificationUser.aspx?Id=12848

[2] Reserve Bank of India. (2025). Reserve Bank of India (Commercial Banks – Credit Facilities) Directions, 2025 (updated July 15, 2026). Reserve Bank of India. https://www.rbi.org.in/Scripts/BS_ViewMasDirections.aspx?id=13156

[3] TransUnion CIBIL & Small Industries Development Bank of India. (2025, May 22). India's MSME commercial credit portfolio grew 13% YoY with overall delinquency rate falling to five-year low [Press release]. SIDBI. https://www.sidbi.in/head/uploads/pressrelease_documents/MSME_Pulse_PR_22_05_2025.pdf

[4] iSPIRT Foundation. (n.d.). Open Credit Enablement Network: Democratising credit. OCEN. https://ocen.dev/

[5] Reserve Bank of India. (2025). Framework for responsible and ethical enablement of artificial intelligence (FREE-AI) in the financial sector. Reserve Bank of India.

[6] Sahamati Foundation. (2026). Data unlocked FY26: Cross-sectoral report on India's Account Aggregator ecosystem. Sahamati. https://sahamati.org.in/

[7] Sahamati Foundation. (2026). Ecosystem dashboard: Account Aggregator adoption metrics. Sahamati. https://sahamati.org.in/aa-dashboard/

[8] Reserve Bank of India. (2026, June 5). Recognition of Sahamati Foundation as Self-Regulatory Organization for the Account Aggregator ecosystem [Press release]. Reserve Bank of India.

[9] ProductNation/iSPIRT. (2023, August 17). Introducing OCEN 4.0: Transforming MSME lending. ProductNation. https://pn.ispirt.in/introducing-ocen-4-0/

[10] NITI Aayog. (2022). India's booming gig and platform economy: Perspectives and recommendations on the future of work. Government of India. https://www.niti.gov.in/sites/default/files/2022-06/25th_June_Final_Report_27062022.pdf

[11] Press Information Bureau, Government of India. (2025, November 17). DPDP Rules, 2025 notified: A citizen-centric framework for privacy protection and responsible data use. Ministry of Electronics and Information Technology. https://static.pib.gov.in/WriteReadData/specificdocs/documents/2025/nov/doc20251117695301.pdf

[12] Ernst & Young. (2026). Decoding the Digital Personal Data Protection Act, 2023 and DPDP Rules, 2025: Compliance guide. EY India. https://www.ey.com/en_in/insights/cybersecurity/decoding-the-digital-personal-data-protection-act-2023

[13] Indian Institute of Banking and Finance. (2025). RBI issues Digital Lending Directions, 2025. IIBF. https://iibf.org.in/documents/Brochure/2025/Digital%20lending%20guidelines%202025.pdf

[14] Suraksha, P. (2024, November 4). GeM Sahay lenders are seeking repayment support from portal. The Economic Times. https://economictimes.indiatimes.com/tech/technology/gem-sahay-lenders-seeking-platforms-repayment-support/articleshow/114918082.cms

[15] TransUnion CIBIL & Small Industries Development Bank of India. (2025). MSME Pulse report, December 2025. SIDBI. https://www.sidbi.in/head/uploads/msmepluse_documents/MSME_Report_8_Dec_2025.pdf

[16] IIFL Finance. (2026, June 25). OCEN business loan: How small shops get working capital at the point of sale. IIFL Finance. https://www.iifl.com/blogs/other/open-credit-enablement-network-ocen-the-future-of-micro-business-credit

[17] AZB & Partners. (2025). RBI (Digital Lending) Directions, 2025 – Same same, but different. AZB & Partners. https://www.azbpartners.com/wp-content/uploads/2025/05/AZB-Update-Digital-Lending-Directions-2025.pdf

[18] Ministry of Labour and Employment, Government of India. (2023). NITI report on gig workers (Rajya Sabha Unstarred Question No. 101). Parliament of India. https://www.sansad.in/getFile/annex/259/AU101.pdf