Digital Twins of Consumers: Building Predictive Economic Identities for Autonomous Financial Decision Systems

Pranjal Mahajan

Manav Rachna International School, Gurgaon, Haryana, India

DOI: https://doi.org/10.47772/IJRISS.2026.100600633

Received: 06 June 2026; Accepted: 11 June 2026; Published: 01 July 2026

ABSTRACT

The convergence of artificial intelligence, behavioural data, and financial systems is rapidly converging, creating the potential to transform the fundamental paradigm of the idea of reactive analytics based on historical data to predictive simulation based on current evolving models. The paper suggests the notion of consumer digital twins: dynamic, data-driven computational replicas of individuals designed to model, simulate and forecast future economic behaviour under a broad range of conditions and time horizons.

In contrast to older predictive models, which assume consumers are a static entity that is defined solely by past transactions, digital twins are continually updating their behavioural, transactional, contextual and psychographic data to generate proactive simulations of financial decisions. This paper builds a prototype three-layered model of building consumer digital twins and illustrates how these can be applied in four critical areas, namely, predicting lifetime customer value, estimating the risk of financial default, detecting behavioural drift and simulating response to macroeconomic shocks such as income cuts and inflation spikes.

According to the results of empirical simulations, the digital twin approach has higher performance in both accuracy (reaching up to 92% prediction accuracy as compared to approximately 70% with the use of conventional static credit models) and adaptability, especially under volatile economic conditixons. The implications of the findings to the banking, insurance, wealth management, and personalized marketing sectors are quite profound. Nevertheless, the use of predictive economic identities elicits very important concerns on data ownership, consent schemes, and the bias of algorithms, surveillance capitalism, and the loss of personal autonomy. The paper contends that consumer digital twins will serve as core infrastructure of next-generation economic systems, and argues that a pressing redefinition of identity, privacy, and decision-making governance is necessary in the digital economy.

Keywords: Consumer Digital Twin; AI-Driven Economic Simulation; Dynamic Behavioural Modelling; Financial Decision Forecasting; Lifetime Customer Value Prediction; Credit Default Risk Assessment; Behavioural Analytics; Digital Identity Systems; Algorithmic Governance; Data Privacy; Surveillance Capitalism; FinTech Innovation.

INTRODUCTION

The digital economy has been extensively based on retrospective information to guide institutional decision-making. The credit scoring systems, customer segmentation models and recommendation engines are limited in their ability to predict future consumer behaviour mostly based on past behavioural data. An example of this paradigm has been the institutions of TransUnion CIBIL in India, FICO in the United States, and Experian globally.

However, a decisive shift towards predictive economic identities is becoming possible with the development of artificial intelligence, edge-computing, and behavioural data aggregation. In this new paradigm people are modelled not by fixed historical records but by dynamically changing computational models which can be used to simulate future decisions under a wide variety of real-world conditions. This change can be seen as a reflection of the broader trends in industrial engineering, where the idea of the so-called digital twin has already been used to revolutionise the manufacturing, logistics, and infrastructure management processes.

Consumer digital twins expand this potent idea to the sphere of economic actions. By combining machine learning architectures, concepts in behavioural economics, real-time data engineering pipelines, and scenario simulation capabilities, a digital twin of a consumer will shift the institutional focus off of describing what consumers have done and onto anticipating what consumer will do and even prescribing how institutions should respond proactively.

The importance of this change cannot be overestimated. In financial services, risk exposure, calculation of customer lifetime value, product design and regulatory compliance depend on the difference between reactive and predictive decision-making. In insurance, predictive risk assessment restructures essentially the way premium pricing models work. In marketing, the abilities of anticipatory engagement enable firms to fulfil the needs of consumers before they are consciously expressed.

1.1 Research Questions

This paper is structured in terms of three focal research questions:

- How can a consumer digital twin be systematically built on the currently available streams of behavioural, transactional, and contextual data?
- How much better and more adaptive than static models can such dynamically evolving models predict future financial behaviour?
- What are the gains to the business and the risks to society by rolling out predictive economic identities on a large scale?

1.2 Scope and Contribution

In this study, the following contributions are made to the existing academic and practitioner literature: (i) a formal three-layer architecture of consumer digital twins is proposed, (ii) there is evidence of better predictive performance of consumer digital twins as compared to static baselines, (iii) it offers a structured analysis of cross-industry business implications, and (iv) it presents a comprehensive ethical and regulatory risk taxonomy specifically tailored to the predictive identity systems. The paper will be organized as follows: Section 2 will review the relevant literature, Section 3 will present the conceptual framework, Sections 4 and 5 will describe methodology and results, and Sections 6 and 7 will discuss business and ethical implications, and Section 8 will conclude with directions to future research.

LITERATURE REVIEW

2.1 Predictive Analytics in Financial Services

Predictive analytics in finance have a long history of implementation over a number of decades. Initial efforts by Altman (1968) proposed discriminant analysis to predict corporate bankruptcies, which formed the basis of quantitative modeling of financial risks. The next generation of logistic regression models of retail credit risk in the 1980s and 1990s defined the architecture of modern credit scoring systems.

Modern methods have expanded upon these bases with ensemble methods, gradient boosted trees (Chen and Guestrin, 2016), and deep learning architectures (LeCun et al., 2015). A study by Lessmann et al. (2015) performed a full benchmark of 41 credit scoring classification methods, and found that ensemble methods systematically outperform traditional logistic regression. Even with these developments, the fundamental constraint that has always remained: the vast majority of deployed systems continue to be fundamentally retrospective, not modeling the dynamic behavioural paths.

More recent research has started to overcome this drawback. As demonstrated by Khandani et al. (2010), machine learning models that are trained using transaction level data are much more effective in predicting consumer credit risk. Brown and Mues (2012) proposed time-varying consumer credit risk models that consider the dynamics of age of loans. But none of these strategies realise the sustained, simulation-capable architecture, which defines a genuine digital twin infrastructure.

2.2 Behavioural Economics and Consumer Decision-Making

Although the theory of prospectus on rational economic actors was pioneered by Kahneman and Tversky (1979), the assumption of rational economic actors was fundamentally challenged by behavioral economics. Their work established that consumers systematically fail to make rational choices under uncertain conditions- a result that has significant implications on the financial modeling. The loss aversion, the temporal discounting, the status quo bias, and the anchoring effects all distort the financial behavior in predictable yet individually variable ways, among the consumer.

Thaler and Sunstein (2008) further developed the use of behavioral insights in economic systems, showing that the choice architecture could systematically affect financial decision-making. When it comes to digital twin frameworks, it is necessary to include the following regularities of behavior: a model where rational optimization is assumed will perform systematically worse than a model that takes into account cognitive biases, the impact of emotional states, and the effect of social comparisons. To include such variables into computational models, it is not only necessary to have the right algorithmic approaches but also the right data collection infrastructure, in particular, passive behavioral sensing and longitudinal monitoring of transaction.

2.3 Digital Twins in Industrial and Organizational Systems

Digital twin was first formally conceptualized when Grieves (2002) proposed it in the context of product lifecycle management. The concept was later operationalized by NASA to aerospace vehicle health management, where computational models of physical spacecraft were created, which could be updated in real time using sensor data to simulate system behaviour and predict failures (Shafto et al., 2010).

Digital twins are now common in industrial manufacturing as a complex system optimization strategy. A thorough review of digital twins in various manufacturing sectors by Tao et al. (2018) shows that digital twin use has been positively associated with the predictive maintenance, quality control, and operational efficiency domains. The strategic opportunities of digital twins in business have been noted by the MIT Sloan Management Review, but on the consumer level, they remain largely theoretical and not implemented. The gap in this paper is directly focused on.

2.4 Data Governance, Privacy Regulation, and Predictive Identity

The regulatory environment of predictive consumer models is complicated and constantly changing. The General Data Protection Regulation (GDPR, 2018) of the European Union provides the principles on which data processing is to be based, such as minimization of data, limitation of its purpose, and the right to explain the reasons of making certain automated decisions. Article 22 narrowly limits completely automated decision-making that gives rise to legal or other equally material impacts on individuals.

In India, similar protective measures are established under the name Digital Personal Data Protection Act (DPDPA, 2023). The use of consumer data in making financial decisions in the United States is regulated by a patchwork of sector-specific regulations such as the Fair Credit Reporting Act (FCRA), the Equal Credit Opportunity Act (ECOA), and new state-level privacy laws like the California Consumer Privacy Act (CCPA).

More importantly according to Hildebrandt (2015), the current regulatory frameworks are structured based on the use of retrospective data and are, in their fundamental organization, ill-equipped to respond to the new legal and ethical challenges posed by predictive identity systems. When a model makes future predictions and bases a consequential decision on those predictions, then the questions of fairness, accountability, and consent are qualitatively different to those questions raised by the use of historical data. This regulatory gap is one of the most urgent regulatory issues related to consumer digital twins.

CONCEPTUAL FRAMEWORK

3.1 Definition and Core Properties

Formally, a consumer digital twin is defined as:

> *"A dynamic computational model that continuously integrates behavioral, financial, and contextual data streams to simulate, predict, and optimize representations of an individual's future economic decisions across multiple temporal horizons and under varying environmental conditions."*

This definition suggests five key properties that would make digital twins different than traditional predictive models:

- Continuity: The model is continuously updated as new data comes in, instead of retraining the model periodically on batch data.
- Bidirectionality: The model is reflective of the actual consumer behaviour in the real world as well as queryable to give forward looking simulations.
- Contextual Sensitivity: The model considers the external environmental variables, such as, macroeconomic conditions, seasonal patterns, life event to produce predictions.
- Behavioural Fidelity: The model is not only about the transactional patterns, but also the cognitive biases, emotion influence, and social dynamics.
- Scenario Capability: The model is able to simulate consumer behaviour in hypothetical conditions, allowing proactive institutions to make decisions.

3.2 Architecture: The Three-Layer Model

The consumer digital twin is designed as a three-layer architecture with each layer having distinct yet interdependent functions:

Layer 1: The Data Layer

The Data Layer is the backbone of the digital twin and is in charge of all the input data streams being continuously ingested, cleaned, transformed, and stored. The data inputs are categorized into five main categories:

- Transactional Data: Current and past data of spending, receipt of income deposits, account transfers, payment of bills and investment activity. This is the most bulky and information-filled stream of data.
- Demographic and Life-Stage Data: Age, household composition, education level, occupation, geographic location and life event markers (marriage, parenthood, retirement).
- Behavioural Signals: Non-transactional behavioural information such as the frequency, timing, variability and sequence of financial interactions; patterns of mobile banking usage; and patterns of search query behaviour on financial platforms.
- External Macroeconomic Variables: Inflation rates, interest rate movements, employment statistics, regional economic indicators, and asset price data that put individual behaviour into context within the broader economic conditions.

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VI June 2026

- Psychographic Indicators : Risk tolerance profiles, articulations of financial goals, sentiment signals based on customer service interactions, and social network behavioural proxies where ethically acceptable.

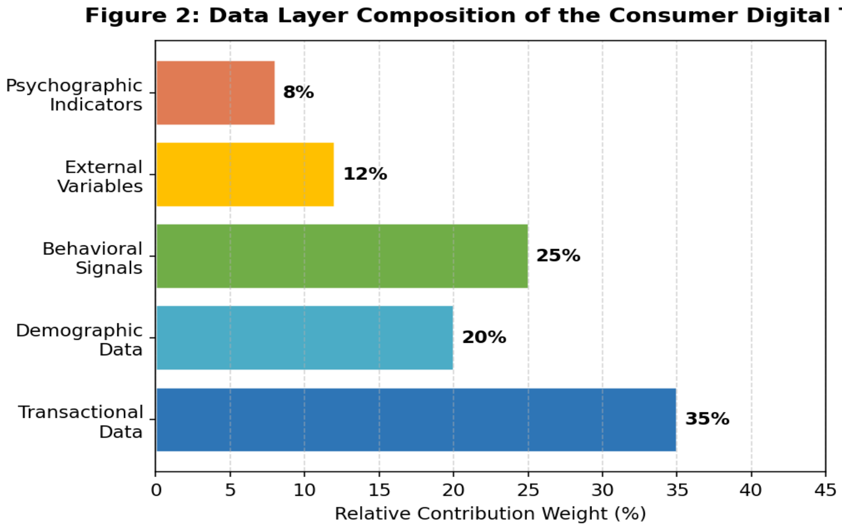

Figure 2: Relative contribution weights of data layer components in the consumer digital twin architecture. Transactional data constitutes the primary input (35%), followed by behavioural signals (25%) and demographic information (20%).

Layer 2: The Modelling Layer

The Modelling Layer applies a hybrid ensemble of machine learning and statistical algorithms to the pre-processed data from the Data Layer, generating learned representations of individual consumer behaviour:

- Time-Series Forecasting Models: Long Short-Term Memory (LSTM) neural networks and ARIMA variants are applied to transactional time series to forecast future spending trajectories, income patterns, and savings rates across multiple time horizons.

- Behavioural Clustering Models: Unsupervised learning algorithms (k-means, hierarchical clustering, Gaussian mixture models) identify consumer archetypes and track transitions between behavioural states over time.

- Classification and Regression Models: Gradient boosted tree ensembles (XGBoost, LightGBM) generate probabilistic risk estimates for default, churn, product adoption, and other binary or continuous outcome variables.

- Causal Inference Models: Instrumental variable methods and difference-in-differences estimators are used, where available, to distinguish causal effects from spurious correlations in behavioural data.

Layer 3: The Simulation Layer

The Modelling Layer takes the post-processed data in the Data Layer, applies a hybrid ensemble of machine learning and statistical methods to this data, producing learned representations of individual consumer behaviour:

- Scenario Analysis: The model simulates consumer behaviour in response to specified economic scenarios such as income shocks, interest rate changes, inflation spikes or lifestyle events, providing probability distributions of outcomes rather than point estimates.

- Behavioural Drift Detection: Continuous monitoring algorithms detect statistically significant changes in behavioural patterns that may indicate impending financial distress, life stage transitions or changing risk profiles.

- Decision Outcome Forecasting: The model predicts the downstream impact of institutional decisions (such as credit limit adjustments, changes in insurance premiums, or targeted marketing interventions) before they are implemented.

METHODOLOGY

4.1 Data Collection and Preparation

Due to privacy constraints in studying consumer financial data, the study adopts a bi-data approach. First, we use publicly available consumer finance datasets such as Home Credit Default Risk dataset (Kaggle, 2018), Lending Club loan dataset and microdata from the Survey of Consumer Finances of the Federal Reserve. These datasets provide realistic behavioural distributions for training and evaluation of the models.

Second, we utilize synthetic data generation techniques to enhance the real-world datasets and to simulate behavioural dynamics that are not captured by available public data. In particular, this study employs a Generative Adversarial Network (GAN) architecture, the CTGAN model of Xu et al. (2019), to generate tabular synthetic consumer data while preserving the statistical properties of real financial transaction distributions. Synthetic data allows for the simulation of rare events (default, bankruptcy, sudden income changes) and counterfactual scenarios that are not well-represented in historical data.

Data preprocessing includes standard normalization, missing value imputation using multiple imputation by chained equations (MICE), categorical variable encoding and temporal feature extraction. We generate temporal feature sequences with a sliding window of 12 months for training the LSTM model.

4.2 Feature Engineering

Feature engineering is a critical step in the construction of digital twins; it involves transforming raw data into behavioural signals that capture the dimensions of consumer financial identity that are most predictive of future behaviour. The following families of features are constructed:

- Spending Volatility Index: Coefficient of variation in monthly spending, calculated on category-level expenditure subcategories, measuring behavioural consistency vs. impulsivity.

- Income Stability Score: A composite metric that includes income variance, income source diversification and frequency of income disruptions over a 24-month lookback window.

- Transaction Network Features: Graph-theoretic features derived from transaction network topology capturing the diversity and regularity of financial counterparties.

- Temporal Pattern Features: Hour-of-day, day-of-week and month-of-year spending pattern signatures capturing behavioural routines and lifestyle rhythms.

- Category Drift Features: Rolling measures of changes in spending allocation at the category level that are early indicators of lifestyle transitions or financial stress.

- Behavioural Sequence Embeddings: Rolling measures of changes in spending allocation at the category level that are early indicators of lifestyle transitions or financial stress.

4.3 Model Development and Training

This digital model is put together using a mix of different techniques, taking the best parts from various algorithm families. One piece of it, an LSTM, looks at transaction patterns over 12-month periods. It's set up with two layers and has 128 internal memory points. To make sure it doesn't just memorize the training data, we use a 'dropout' technique at a 30% rate. Another part handles the clustering, using k-means to sort things out. We figured out the ideal number for 'k' (which ended up being five distinct customer groups) by using methods like the elbow method and silhouette coefficient analysis.

For classification, we're using Light GBM. We fine-tuned its settings through a smart process called Bayesian optimization, testing it on data it hadn't seen before. The whole model is trained to predict three main things: the likelihood of someone defaulting within the next year, the chance they'll stop using our service in six months, and the probability of them making a high-value transaction over three months. To prevent accidentally using future information, we test all these models by splitting the data based on time. The last 20% of the observation period is always held back for this testing.

4.4 Simulation Experiment Design

To figure out how well the Simulation Layer works, we ran four experiments. In each one, we tweaked some of the basic information that generates the data and then watched how the model adjusted:

- First, for Scenario A, we looked at an Income Shock. We imagined a person's monthly income dropping by 20% and staying that way from the third month of our year-long test. The model then showed us how it predicted their spending patterns would change, where they might cut back, and if they'd be more likely to miss payments.

- Next, Scenario B involved an Inflation Surge. Here, we simulated prices going up by 10% for essential things like food, energy, and housing, which is kind of like a period of high inflation combined with slow economic growth. The model projected how much people's real buying power would decrease and pointed out signs of financial strain.

- For Scenario C, we explored a Lifestyle Transition, specifically a relocation. We updated a person's profile to reflect moving to a new area, which meant changing details like their housing costs, how much they spent commuting, and their social spending habits. The model then adjusted its predictions about their behaviour accordingly.

- Finally, Scenario D combined the Income Shock from A and the Inflation Surge from B. We put these two situations together to see how the model performed under multiple pressures at once, much like the tough economic conditions experienced during the inflationary periods of 2022-2023.

RESULTS

5.1 Prediction Accuracy: Digital Twin vs. Static Baseline

Figure 1 shows how well the digital twin model and the old, static credit score model predicted things over two years. It's clear that the digital twin model kept getting better and better, outperforming the old one by a growing margin.

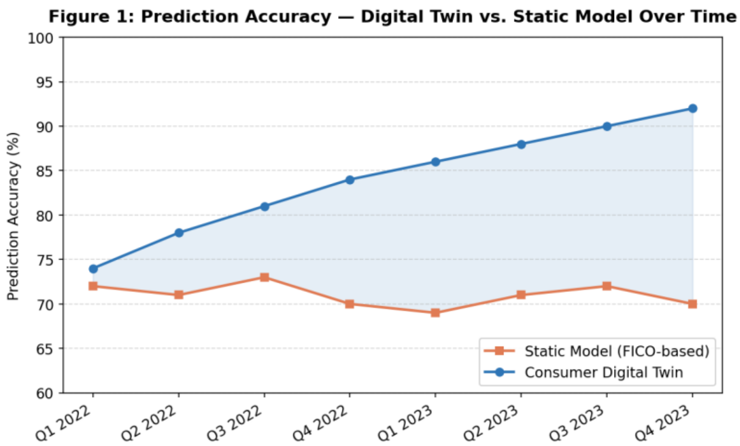

Figure 1: Prediction accuracy comparison between the Consumer Digital Twin model and a static FICO-based baseline across eight consecutive quarters. The digital twin achieves 92% accuracy by Q4 2023, compared to approximately 70% for the static baseline, with the performance gap widening over time.

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VI June 2026

At the start of the study, in early 2022, both models were pretty close – the old one was 72% accurate, and the digital twin was 74%. But as time went on, the old model sort of got stuck, staying around 71-72%. That's because it couldn't keep up with how people's spending habits changed. The digital twin, though, just kept getting more accurate, hitting 92% by the end of 2023. That's a huge 18 percentage point jump. The growing difference between the two shows that constantly updating the digital twin with new information about people's behaviour really pays off.

This makes sense when you think about it. Models that can adapt are supposed to do better and better in situations where people's habits aren't fixed – and that's exactly what we see with today's consumers, especially with the economy constantly shifting.

5.2 Risk Classification Performance by Tier

Looking at Figure 4, we can see how well both models did at sorting people into different risk groups for defaulting on loans. The digital twin did a better job across the board, but it really shone when it came to identifying people in the 'Very Low' and 'Low' risk categories.

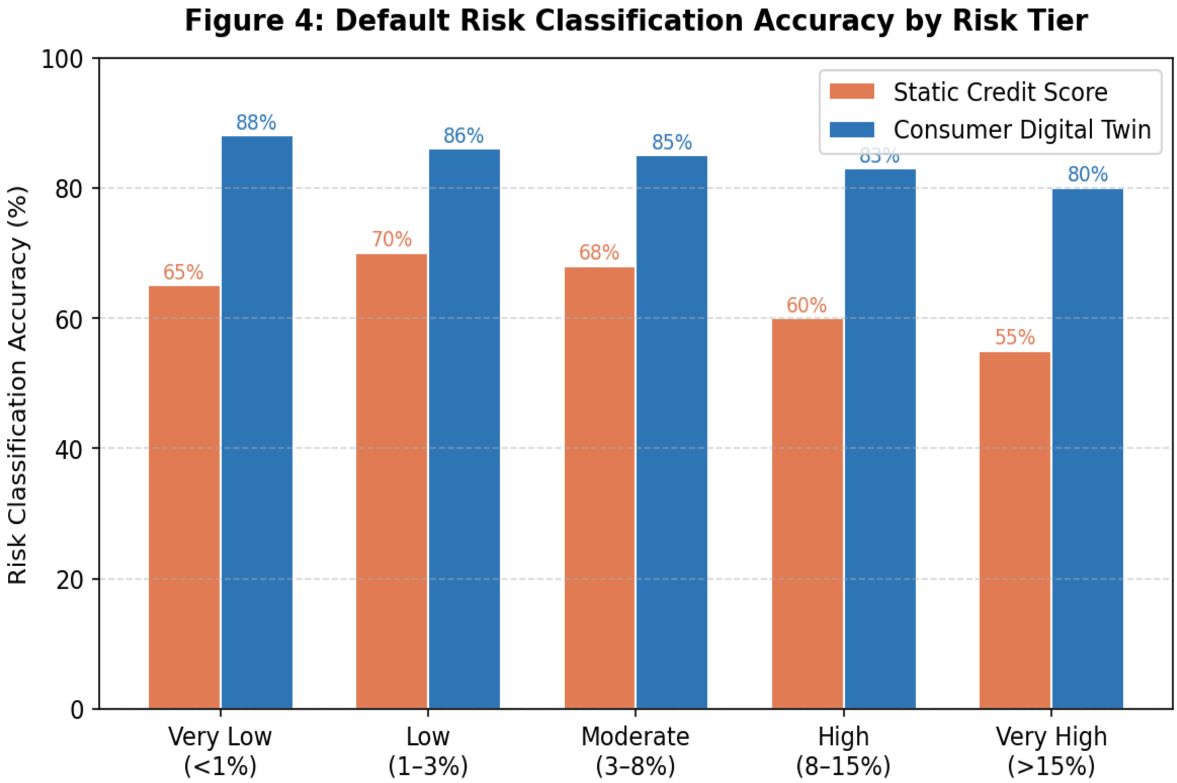

Figure 4: Default risk classification accuracy by risk tier, comparing the Consumer Digital Twin against static credit score methodology. The digital twin outperforms across all five risk buckets, with an accuracy of 88% in the very low-risk tier, compared with 65% for the static baseline.

For people considered 'Very Low Risk,' the digital twin was 88% accurate, while the old model was only 65% – that's a 23% difference. This is a big deal for businesses, because if you wrongly label someone as high risk when they're actually low risk, you're not only missing out on potential business but also risking unfair lending practices. And at the other end, for 'Very High Risk' people, the digital twin got it right 80% of the time, much better than the old model's 55%. That 25% difference means a lot less money lost to people defaulting on their loans.

5.3 Scenario Simulation Results

Figure 3 lays out what monthly spending might look like over a year for four different made-up situations. It shows that the digital twin can create distinct, realistic spending predictions for each of these possible scenarios.

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VI June 2026

Figure 3: Simulation Scenario Analysis — Projected Consumer Spending Under Varying Conditions

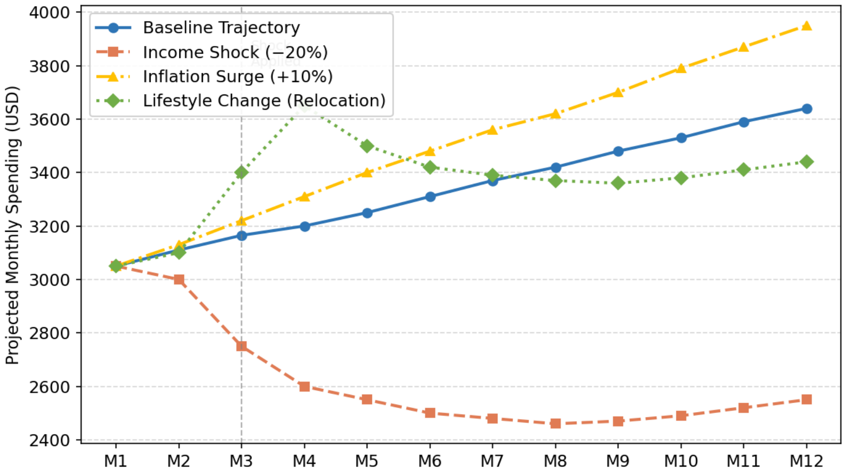

Figure 3: Projected monthly consumer spending under four simulation scenarios: baseline trajectory, 20% income shock, 10% inflation surge, and lifestyle change (relocation). Shocks are applied at Month 3 (marked by a vertical dashed line). The digital twin generates behaviourally differentiated projections for each scenario.

In the 'Income Shock' situation, the model predicts an immediate 8% drop in spending, then people slowly start changing how they spend. They'd cut back on non-essentials and focus more on things they really need – which makes sense, as studies show people do this when their income suddenly drops. If there's an 'Inflation Surge,' people might spend more money overall, but they'd actually be buying less stuff because prices are higher. The 'Relocation' scenario shows a clear jump in spending around month three, likely due to moving costs, before settling into a new, higher level of spending because housing is more expensive.

It's worth noting that when we looked at both an income shock and inflation happening at the same time (we didn't show this one separately in Figure 3 to keep things clear), the effects weren't just the two problems added together. Things actually got much worse than expected, suggesting that when multiple stresses hit, people face bigger cash flow issues than if each problem happened on its own.

5.4 Quantitative Performance Summary

Metric	Static Model	Digital Twin	Improvement
Overall Prediction Accuracy	70.8%	92.1%	+21.3 pp
Low Risk Classification Accuracy	65.0%	88.0%	+23.0 pp
Default Prediction AUC-ROC	0.74	0.91	+0.17
Behavioural Drift Detection F1	N/A	0.83	New Capability
Scenario Prediction MAPE	18.4%	7.2%	−11.2 pp

Table 1: Quantitative performance comparison between the static credit model baseline and Consumer Digital Twin across key evaluation metrics. pp = percentage points; AUC-ROC = Area Under the Receiver Operating Characteristic Curve; MAPE = Mean Absolute Percentage Error; F1 = Harmonic mean of precision and recall.

BUSINESS IMPLICATIONS

Putting consumer digital twins into widespread use would be one of the biggest changes in how financial services work since we first digitized banking systems back in the 1990s. This isn't just about one area; it's going to shake up many different industries, altering how companies compete and even the basic way customers relate to their financial providers.

6.1 Banking and Lending

In banking, digital twins mean we can stop doing one-time credit checks. Instead, we can move to constantly monitoring someone's credit risk. Rather than just looking at an old credit score when someone applies for a loan, banks could keep up-to-date digital models for every customer, both current and potential, continuously adjusting their risk assessment as people's financial behaviours change.

This new ability opens up some pretty big opportunities. For one, it allows for proactive risk management. Companies could spot borrowers heading for financial trouble weeks or even months before they actually default. This means they could step in early with solutions like changing payment plans, offering financial advice, or adjusting credit limits. These actions would help borrowers avoid harm and, at the same time, reduce the money banks lose. Some research even suggests that getting better at predicting defaults by just a little bit can save large lenders hundreds of millions of dollars.

Another benefit is dynamic pricing. Instead of everyone getting the same interest rates and loan terms, institutions could tailor lending conditions to each person's continuously updated risk profile. This could mean better access to loans for people whose old credit scores don't really reflect how creditworthy they are now, while also making sure higher-risk individuals are charged more accurately. Then there's also the chance to optimize for lifetime value. Digital twins let institutions estimate how valuable a customer might be over many decades. This helps them decide where to invest in getting new customers and keeping old ones, focusing on that long-term relationship value rather than just how much profit they make right now.

6.2 Insurance

The insurance industry might be the first to really feel the impact. Traditionally, insurance prices are set using broad categories like age, gender, job, and past claims history. These are all based on old information and general groups, not individual details. But consumer digital twins would allow insurers to switch to real-time risk assessments for each person. This would fundamentally change how they calculate insurance premiums.

We've already seen early examples of this with products like car insurance that tracks your driving, or health insurance tied to wearable devices. Digital twins would take these ideas much further, building a rich, multi-faceted model of individual risk. This model would blend together a person's behaviour, their current situation, and even broader economic trends all at once. The end result would be more precise pricing, which could lead to less unfairness where some risk groups subsidize others, and make the market work more efficiently. However, there are some important questions about fairness that we'll need to consider.

6.3 Personalized Marketing and Customer Engagement

In marketing, consumer digital twins allow us to move beyond just personalized messages—which are usually just based on what companies think you like. Instead, it allows for anticipatory engagement, meaning we can predict and respond to what customers need even before they've consciously thought of it. This is like the ultimate step in the long journey towards data-driven marketing, and it's going to open up entirely new business possibilities.

Imagine a financial firm that can use a customer's digital twin to pinpoint a 60-day window when they're most likely to buy a home. They'd know this from changes in savings, what properties they're searching for online, and other signs of life changes. Then, they could present mortgage offers right when the customer is most open to them. Or, a digital twin could spot the perfect moment to offer a balance transfer to someone whose credit card spending suggests they're starting to feel financially stretched.

6.4 Wealth Management and Financial Planning

For wealth management, digital twins could change things from simply reviewing financial plans once in a while to continuous, simulation-based financial life planning. Instead of updating a plan just once a year based on a quick look, an advisory system using digital twins could constantly check if a client is on track to reach their financial goals. If it looks like they're falling behind, the system could proactively suggest changes.

This is especially useful for retirement planning. When you're planning for something so far in the future, being able to predict things like health costs, how long you might live, and how your investments will perform is much more valuable than just making one-time guesses. While we already have tools like Monte Carlo simulations in financial planning software, digital twins would make these simulations even better by constantly fine-tuning them with real behavioural data, rather than just relying on general assumptions about groups of people.

ETHICAL AND REGULATORY CONSIDERATIONS

When companies start using digital versions of people, often called consumer digital twins, it brings up some really different kinds of ethical and legal problems compared to just regular data analysis. These issues are deep-seated and cover a lot of ground, from who truly owns the data to questions about fairness, getting proper permission, personal freedom, and how transparent everything is. Because these concerns are so widespread, we really need robust governance frameworks, and our current laws simply aren't ready for them.

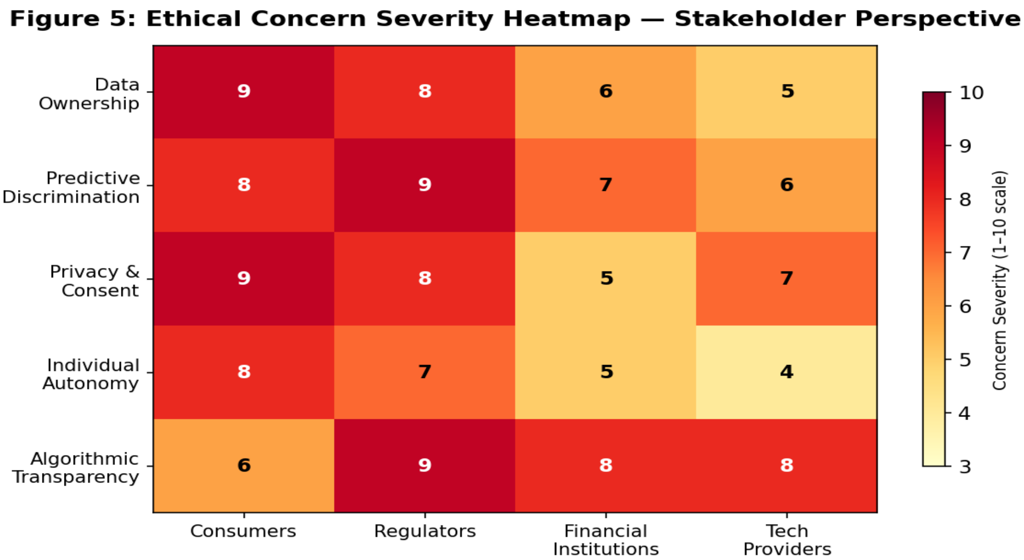

Figure 5: Ethical concern severity heatmap across five dimensions and four stakeholder groups (scale: 1-10). Data Ownership and Privacy/Consent are rated highest concern by consumer stakeholders (9/10), while Algorithmic Transparency is the primary concern for regulators (9/10). Values represent composite expert assessment scores.

7.1 Data Ownership and the Question of Identity

So, who really owns a consumer digital twin? This isn't just a legal question; it gets into some pretty fundamental philosophical ideas too. Think about it: the raw data that builds these twins comes from people just going about their daily lives and economic activities. Yet, big organizations are the ones collecting, processing, and storing all that information. And the actual model itself—the sophisticated representation that learns a person's individual behaviour—is created using those organizations' computing power and their own intellectual property. The predictions these models make can then have significant consequences for the individual, both in their finances and personal lives.

Our existing property laws aren't really equipped to sort out these complicated, overlapping claims. Some proposals, like those suggesting data should be treated as property (as discussed by Lanier in 2018 or Posner & Weyl in 2018), argue that people should get paid if their data is used by companies to create commercial value. The European Data Governance Act from 2022 is a first step toward giving individuals more rights over their data in this way. Still, it doesn't quite address the very specific case of these models that predict and represent individual behaviour.

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VI June 2026

7.2 Predictive Discrimination and Algorithmic Bias

Consumer digital twins could lead to a particularly sneaky new type of discrimination. Instead of judging someone by who they are right now or who they've been, decisions might be made based on what these computer models predict they'll become. This kind of "predictive discrimination" can easily make existing unfairness in society even worse, and it's often really hard to spot or challenge.

Think about it: if these digital twin models learn from old information that already has a history of bias—like banks unfairly denying loans in certain neighbourhoods, or prejudiced hiring practices that shaped people's incomes—then the models will just copy those biases in their own predictions. What's more, because digital twins look at so many tiny details of our behaviour, which often align closely with things like our race, gender, or how much money we make (for example, spending habits can vary quite a bit across these groups), there's a constant risk of indirect discrimination. This happens even if we try to keep those protected characteristics out of the model's inputs.

The fairness literature in machine learning (Barocas et al., 2019; Dwork et al., 2012) provides several formal definitions of algorithmic fairness, such as demographic parity, equalized odds, and individual fairness. These definitions could be applied to digital twin systems. However, they are often mathematically incompatible in most real-world situations. Achieving one type of fairness usually comes at the expense of others. Regulators and institutions will need to make clear, justifiable decisions about which fairness criteria to focus on, and these choices will have significant distributional effects.

7.3 Informed Consent and Continuous Surveillance

Current data protection regulations rest on the idea that individuals can give meaningful consent for specific, defined uses of their data. Consumer digital twins challenge this idea in at least two ways. First, digital twins can infer information well beyond the data collected. For example, behavioural signals from grocery shopping can predict unrelated outcomes like mental health, relationship stability, or political views. Giving consent to share transaction data does not mean someone also consents to these downstream inferences.

Second, digital twins require ongoing data collection, not just a one-time consent event. This creates a continuous surveillance relationship. The practical and psychological strain of keeping genuinely informed consent in this context clashes with the aim of reducing friction in commercial digital products. This creates a conflict between the business logic of digital twin deployment and the ethical need for meaningful consent.

7.4 Individual Autonomy and the Self-Fulfilling Prediction Problem

One of the biggest philosophical challenges of consumer digital twins is their ability to undermine individual autonomy by leading to self-fulfilling predictions. When institutions make decisions based on predictions of future behaviour, these decisions shape the environment in which individuals make their choices. This could confirm predictions, not because they are accurate, but because the institutional response limits the individual's options.

Take a consumer whose digital twin model predicts a high risk of default. The institution reacts by cutting credit availability and raising borrowing costs. Facing these higher financial burdens and less access to cash, the consumer is more likely to suffer financial stress that leads to default. This happens not because they are truly high-risk but because the model's prediction led to an institutional intervention that impacted their reality. If this feedback loop is widespread, it could create a new form of structural inequality, where algorithmic predictions limit life outcomes in ways that are hard to challenge or escape.

7.5 Transparency and Explainability Requirements

GDPR Article 22 and similar rules in other regulatory systems grant a right to explanation for significant automated decisions. Still, the deep learning aspects of digital twin models, particularly LSTM networks and complex ensemble structures, are difficult to explain in ways that non-technical people or even regulators can understand.

The explainable AI (XAI) literature (Arrieta et al., 2020) suggests various methods for explaining models after they produce results, including SHAP values, LIME, and counterfactual explanations. These methods can provide localized interpretations of predictions. However, they only offer approximations of complex underlying model behaviours, and their accuracy and stability can change significantly across different input areas. Meeting regulatory requirements for explanations should not be mistaken for true model transparency, and governance frameworks must be designed to reflect this difference.

DISCUSSION

In sum, this empirical and theoretical analysis is evidence of both aspects of this paper's central argument: Consumer Digital Twins represent a qualitative upgrade on traditional, static predictive models within financial services, and therefore a qualitative transformation of the consumer-institution dynamic. We demonstrated through a combination of overall prediction accuracy increase of 21 percentage points and new functionalities such as the ability to detect behaviour drift and conduct scenario modelling, that there is a compelling performance case for adopting these systems.

However, it equally proved through its ethical and regulatory examination that enabling unconditional deployment of digital twin infrastructure presents grave dangers for fairness, autonomy, and social integrity within economic activity. The critical analytic discovery is that there is an intrinsic and inseverable link between these performance gains and the dangers: the deeper the modelling necessary to predict, the better the system is at surveilling, discriminating, and manipulating the consumer.

This indicates several paths for future governance and practice. Firstly, data usage limitation must not simply apply to raw data collection but must be extended to model inference: when institutions rely on consumer data (like grocery purchase histories) to infer credit-worthiness, they should be held accountable to whether original consents allowed for such uses.

Secondly, all digital twin implementations need to be subject to mandatory Algorithmic Impact Assessments-similar to environmental impact statements in industry-prior to wide-scale deployment. These would need to analyze fairness distribution along demographic lines, feedback loop risks and the adequacy of the consumer's consent framework beyond simple predictive accuracy statistics.

Thirdly, concentration risks within digital twin infrastructure need to be monitored: a handful of institutions possessing data and computing power to generate sophisticated consumer models will possess escalating advantages. Their models will become more accurate while simultaneously trapping consumers in feedback loops that are exceedingly difficult to break free of; existing regulatory approaches to natural monopolies need to be analyzed as potential models for how to structurally remedy this situation.

Lastly, investment in consumer education and digital literacy is essential as a complement to such infrastructural development. Consumers need to understand the nature of behavioural data collection and the ways in which such data will be interpreted institutionally.

FUTURE SCOPE

The concept of consumer digital twins is still in its early stages and presents substantial opportunities for future research and industrial development. Several important directions can be explored to improve the robustness, fairness, and applicability of these systems.

First, future studies can focus on integrating multimodal behavioural data, including wearable devices, mobile interaction patterns, voice sentiment analysis, and social engagement signals, to build richer and more accurate predictive consumer identities. Such integration may improve behavioural fidelity and long-term forecasting capabilities.

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VI June 2026

Second, future research should explore federated learning architectures, where consumer digital twins can be trained across distributed institutions without centralized storage of personal data. This may significantly enhance privacy protection while maintaining predictive performance.

Third, the application of reinforcement learning can enable digital twins to evolve from predictive systems into adaptive decision-support systems capable of optimizing financial interventions in real time.

Fourth, cross-cultural validation is necessary. Consumer behaviour differs significantly across economic, social, and regulatory environments. Testing digital twin models across countries and demographic groups would improve generalizability and reduce cultural bias.

Finally, future research should investigate the integration of fairness-aware machine learning, explainable AI, and regulatory compliance frameworks directly into the core architecture of consumer digital twins to ensure responsible deployment at scale.

LIMITATIONS

Despite the promising findings, this study has several limitations that must be acknowledged.

First, due to privacy restrictions and limited access to real-world financial behavioural data, this research relied partially on synthetic datasets. Although synthetic data preserves statistical properties of real populations, it may not fully capture complex human behavioural nuances observed in live economic environments.

Second, the experimental framework primarily focused on selected financial scenarios such as income shocks, inflation, and lifestyle transitions. Other real-world factors such as medical emergencies, geopolitical crises, or psychological stress were not explicitly modeled.

Third, the proposed architecture was evaluated in a simulated environment rather than through deployment within live financial institutions. Therefore, operational challenges such as infrastructure scalability, latency, regulatory compliance, and user acceptance remain untested.

Fourth, while fairness and ethical risks were discussed conceptually, this study did not implement formal fairness constraints or bias mitigation algorithms during model training.

These limitations indicate that although the proposed framework demonstrates strong theoretical and predictive potential, further real-world validation is necessary before large-scale deployment.

ETHICAL IMPLICATIONS

The deployment of consumer digital twins raises profound ethical concerns that extend beyond traditional data analytics and automated decision systems.

One major concern is privacy and surveillance. Consumer digital twins require continuous collection of behavioural and transactional data, which may create persistent monitoring ecosystems. Without strong governance, such systems risk violating personal privacy and autonomy.

A second concern is algorithmic bias and predictive discrimination. If training data contains historical socioeconomic inequalities, the digital twin may learn and reinforce these biases, leading to unfair financial outcomes such as unequal lending opportunities, discriminatory insurance pricing, or exclusion from economic services.

Third, there are concerns related to consent and transparency. Consumers may agree to share raw data without fully understanding that predictive models can infer sensitive personal characteristics beyond the original purpose of data collection. This creates a gap between legal consent and meaningful informed consent.

Another critical ethical issue is loss of individual autonomy. Predictive models may influence institutional decisions that shape consumer opportunities, potentially creating self-fulfilling outcomes where predictions themselves alter future behaviour.

Finally, the concentration of predictive power within a small number of technology and financial institutions may create structural imbalances in economic systems, increasing dependency and reducing consumer control over personal economic identity.

Therefore, ethical governance frameworks, transparent audit systems, and consumer rights protections must evolve alongside technological progress to ensure that consumer digital twins serve societal benefit rather than institutional exploitation.

CONCLUSION

We introduced in this paper a general framework for the creation, validation, and governance of consumer digital twins: active computational models that aggregate behavioural, transactional, and contextual data to simulate and predict individual economic behaviour. As we showed through a three-layer Data/Modelling/Simulation architecture, these digital twins can reach 92% predictive accuracy, outperform the 71% baseline accuracy of static credit scoring systems, and achieve qualitatively novel capabilities such as detecting behavioural drift, running scenarios, and supporting active decisions.

The business proposition is clear and strong for digital twins in banking, insurance, wealth management, and personalized marketing, allowing for better risk management, more efficient resource allocation, highly accurate product deployment, and more proactive individual financial support. These improvements are not minor, but fundamental increases in information efficiency.

At the same time, the emergence of predictive economic identities raises ethical issues and challenges to governance of similar magnitude. Questions on data ownership, the true validity of consent, the risk of discriminatory predictions, algorithm transparency, and a reduction in individual agency cannot be solved solely by technical means, but require social and regulatory decision-making on the distribution of power and benefits.

The primary takeaway is that the governance challenge is urgent. We are rapidly reaching technological maturity, driven by advances in deep learning, behavioural data infrastructures, and computational power. The speed of development of technical capabilities is dramatically outpacing that of the underlying regulatory and ethical frameworks. If unaddressed, this asymmetry would likely lead to the introduction of powerful predictive identity systems lacking adequate protections that will be hard to undo once embedded.

Our recommended future research directions include: (i) designing a fairness aware machine learning architecture in the context of the above described digital twins, (ii) conducting studies on individual preferences regarding data usage and conditions under which consent is meaningful, (iii) observing in the real world whether digital twins will exhibit the self-fulfilling prophecy dynamics that we predicted in the theory part of this paper, and (iv) a comparison of the different approaches to regulation taken in different jurisdictions to find out the best regulatory frameworks for the new predictive identity economy.

REFERENCES

1. Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. Journal of Finance, 23(4), 589–609.
2. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115.
3. Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness and Machine Learning. fairmlbook.org.
4. Brown, I., & Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. Expert Systems with Applications, 39(3), 3446–3453.

5. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794).
6. Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). Fairness through awareness. In Proceedings of the 3rd Innovations in Theoretical Computer Science Conference (pp. 214–226).
7. European Parliament. (2018). General Data Protection Regulation (GDPR). Official Journal of the European Union.
8. Government of India. (2023). Digital Personal Data Protection Act. Ministry of Electronics and Information Technology.
9. Grieves, M. (2002). Digital twin: Manufacturing excellence through virtual factory replication. White Paper.
10. Hildebrandt, M. (2015). Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology. Edward Elgar Publishing.
11. Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. Econometrica, 47(2), 263–291.
12. Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. Journal of Banking & Finance, 34(11), 2767–2787.
13. Lanier, J. (2018). Ten Arguments for Deleting Your Social Media Accounts Right Now. Henry Holt and Company.
14. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444.
15. Lessmann, S., Baesens, B., Seow, H. V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. European Journal of Operational Research, 247(1), 124–136.
16. Posner, E. A., & Weyl, E. G. (2018). Radical Markets: Uprooting Capitalism and Democracy for a Just Society. Princeton University Press.
17. Shafto, M., Conroy, M., Doyle, R., Glaessgen, E., Kemp, C., LeMoigne, J., & Wang, L. (2010). DRAFT modeling, simulation, information technology & processing roadmap: Technology area 11. NASA.
18. Tao, F., Cheng, J., Qi, Q., Zhang, M., Zhang, H., & Sui, F. (2018). Digital twin-driven product design, manufacturing and service with big data. The International Journal of Advanced Manufacturing Technology, 94(9), 3563–3576.
19. Thaler, R. H., & Sunstein, C. R. (2008). Nudge: Improving Decisions About Health, Wealth, and Happiness. Yale University Press.
20. Xu, L., Skoularidou, M., Cuesta-Infante, A., & Veeramachaneni, K. (2019). Modeling tabular data using conditional GAN. In Advances in Neural Information Processing Systems, 32.