FinTech Lending in India: Taking Stock of Implications for Privacy and Autonomy

Vidushi Marda and Amber Sinha*

ABSTRACT *In the last five years, the Fintech sector has thrived in India, with Machine Learning (ML) driven credit scoring based on alternative data, emerging as a growing segment. The credit scoring industry in India needs to be viewed in light of a careful examination of rights, inclusion, appropriate safeguards and discrimination, currently missing from the discourse and practices. In this paper, we explain how ML-based credit scoring works, and the regulatory and commercial factors that have enabled and impeded its growth in India. Through legal and technological analysis, richened by insights from qualitative interviews with entrepreneurs and practitioners, we provide a detailed picture of the credit scoring sector, and highlight its spillover privacy and predatory impacts in India.*

I. INTRODUCTION

Financial Technology ("FinTech") is defined as the intersection of the financial services and technology sectors, where technology-focused start-ups and new market entrants innovate the products and services traditionally provided by the financial services industry.[1] There are over seven thousand

---

*   Vidushi is a Lawyer and Senior Programme Officer at Article 19. Amber is a Lawyer and Trustworthy AI-Senior Fellow at Mozilla Foundation.

FinTech companies in India, next only to the US and China.[2] In 2021, there were investments of over USD 8 million in this segment in India.[3] This includes a wide range of financial services, from lending and payments, to alternate credit scoring and insurance. There are several factors that led to the rise of the FinTech industry in India. The Indian government's investment in a digital ecosystem, its focus on increasing bank accounts accessible through Aadhaar-based verification, and the decision to demonetise about 85% of currency in circulation in 2016 have all contributed to a boost in the Fintech sector. The sector promises nimbler and cost-effective financial services, and can enable financial inclusion through a range of services including new methods of risk assessment and mobile wallets. The potential for India's FinTech sector to attract global investment and incentivise economic growth has also attracted attention from start-ups, investors,[4] and regulators.[5] Further, these services have evolved in the context of the narrative of financial inclusion—over 95% of the Indian population has no credit history as they lack the financial truncation history to generate CIBIL scores.[6] The Government's push towards a digitally empowered society is visible through its Digital India initiative.[7] The Reserve Bank of India has played a key role in enabling FinTech companies to emerge and operate in India by encouraging innovation and providing the regulatory and infrastructural capabilities to

---

[1]   'Blurred Lines: How FinTech is Shaping Financial Services' (*PwC*, March 2016) <https://www.pwc.de/de/newsletter/finanzdienstleistung/assets/insurance-inside-ausgabe-4-maerz-2016.pdf> accessed 23 April 2023.

[2]   'At $29 bn, Indian Fintech Sector Now has 14% Global Funding Share: Report' *Business Standard* (New Delhi, 22 August 2022) <https://www.business-standard.com/article/companies/at-29-bn-indian-fintech-sector-now-has-14-global-funding-share-report-122082201014_1.html> accessed 23 April 2023.

[3]   Naina Bhardwaj, 'India Briefing, What Trends are Driving the Fintech Revolution in India?' (*India Briefing*, 9 June 2022) <https://www.india-briefing.com/news/what-trends-are-driving-the-fintech-revolution-in-india-23809.html/> accessed 23 April 2023.

[4]   Arti Singh, 'Fintech VC Report Card— Part III: Omidyar vs. Kalaari vs. Blume vs. Prime vs. Ribbit' *The Economic Times* (29 January 2019) <https://prime.economictimes.indiatimes.com/news/67733067/fintech-and-bfsi/fintech-vc-report-card-part-iii-omidyar-vs-kalaari-vs-blume-vs-prime-vs-ribbit> accessed 23 April 2023.

[5]   'Initiatives by India's Government to Boost FinTech' (*FinTech Futures*, 2 January 2019) <https://www.fintechfutures.com/2019/01/initiatives-by-indias-government-to-boost-fintech/> accessed 23 April 2023.

[6]   Tarunima Prabhakar, CLTC White Paper Series, A New Era for Credit Scoring: Financial Inclusion, Data Security, and Privacy Protection in the Age of Digital Lending (*Centre for Long-Term Cybersecurity, University of California Berkeley,* June 2020). <https://cltc.berkeley.edu/wp-content/uploads/2020/06/A_New_Era_for_Credit_Scoring.pdf> accessed 23 April 2023.

[7]   'Digital India - A Programme to Transform India into Digital Empowered Society and Knowledge Economy' (*Press Information Bureau- Government of India*, 20 August 2014) <http://pib.nic.in/newsite/PrintRelease.aspx?relid=108926> accessed 23 April 2023.

do so.[8] Further, the Artificial Intelligence Task Force set up by the Ministry of Commerce and Industry identifies FinTech as a domain of relevance and priority for the Government of India.[9]

The digital lending industry in India has benefited from government and regulatory support and grew rapidly chiefly due to two reasons.[10] First, Aadhaar, the biometric identity scheme in India, meant that companies could verify and onboard potential lendees at virtually no cost in terms of time and money. Second, the ability to scrape intimate details about a person's life from social media, text messages, call records etc. using sophisticated algorithmic and statistical models meant that analysis of profiles could be much more granular at a negligible cost.

In 2018, there were two significant changes in the ability of FinTech companies to take advantage of these options. First, following the Supreme Court's verdict on Aadhaar, the ability to use Aadhaar numbers for onboarding customers has been significantly curtailed. Second, the Personal Data Protection Bill was introduced, revised several times and finally withdrawn. The draft of a new Digital Data Protection Bill has been released by MeitY and 2022, is yet to be passed, with significant implications for the use and processing of data. Therefore, despite the Supreme Court's decisions on right to privacy and restrictions on private use of Aadhaar, the Fintech industry has grown in India, with little, if any regulation of the data ecosystem that it relies on. This recent development in the absence of data governance provisions had direct and clear implications for the privacy and autonomy of individuals who are the primary customers of this industry.

Against this background, this paper will study the impact of these changes on the FinTech lending sector in India and subsequent developments, with specific reference to implications for privacy and autonomy. It aims to do so to bridge some gaps between academic analysis and industry insights in the context of alternate lending. Section I provides background and an introduction to this report. Section II will offer a primer on the privacy and security opportunities, limitations, and vulnerabilities offered by the two

---

8   Report of the Working Group on FinTech and Digital Banking (*Reserve Bank of India*, 2017) <https://rbidocs.rbi.org.in/rdocs/PublicationReport/Pdfs/WGFR68AA1890D7334D8F8F72CC2399A27F4A.PDF> accessed 23 April 2023.

9   Report of the Artificial Intelligence Task Force (*Department for Promotion of Industry and Internal Trade*, 20 March 2018) <https://dipp.gov.in/whats-new/report-task-force-artificial-intelligence> accessed 23 April 2023.

10 Gopal Sathe 'After Beta-Testing on a Billion Indians, The Tech behind Aadhaar is Going Global: Modi Bats for India Stack at Singapore Summit' (*HuffPost India*, 12 June 2018) <https://www.huffingtonpost.in/2018/06/06/after-beta-testing-on-a-billion-indians-the-tech-behind-aadhaar-is-going-global_a_23452248/> accessed 23 April 2023.

technologies that form this report's focus: Aadhaar-based authentication, and machine-learning based lending in the FinTech sector. Section III will analyze the policy developments that have had an impact on FinTech companies' ability to conduct business, and lay out the current state of affairs. Section IV will contextualize analysis with perspectives from practitioners in the FinTech sector, gathered through a series of qualitative interviews. Section V will conclude with findings and recommendations.

II. EVALUATING UNDERLYING TECHNOLOGIES

A. Aadhaar-based authentication

Aadhaar, the largest biometric identity project in the world, was introduced in 2009 by the Government of India. It intends to provide unique identification for Indian residents that can be used for the efficient delivery of services. The Unique Identification Authority of India (UIDAI) is the authority in charge of Aadhaar enrollment and authentication, created to issue unique identification numbers that are *"(a) robust enough to eliminate duplicate and fake identities, and (b) can be verified and authenticated in an easy, cost-effective way."*[11] At the time of enrollment of individuals into the Aadhaar system, both biometric and demographic details such as name, date of birth, and address are collected. This includes fingerprints, iris scans, and photographs of each individual being enrolled. This data is stored in the Central Identities Data Repository ("CIDR").

A key component of FinTech lending is the process of Know Your Customer ("KYC") - the due-diligence that lenders need to carry out at the time of verifying and assessing potential customers by obtaining appropriate information about them.[12] Following the Aadhaar Act of 2016, private companies were allowed to use Aadhaar - which meant that lenders could lower compliance costs to carry out KYC and customer onboarding, essentially completing the process in a matter of minutes as opposed to a few days.[13]

---

[11] 'About UIDAI' (UIDAI - Government of India) <https://uidai.gov.in/en/about-uidai/unique-identification-authority-of-india.html#:~:text=The%20Unique%20Identification%20Authority%20of,the%20Ministry%20of%20Electronics%20and> accessed 23 April 2023.

[12] Reserve Bank of India, 'Guidelines on Digital Lending' (*Reserve Bank of India*, 2 September 2022) <Lending. https://rbidocs.rbi.org.in/rdocs/notification/PDFs/GUIDE LINESDIGITALLENDINGD5C35A71D8124A0E92AEB940A7D25BB3.PDF> accessed 23 April 2023.

[13] KYC Solutions, 'Problems and Challenges in Traditional KYC Systems' (*Records Keeper,* December 2016) <https://www.recordskeeper.com/blog/kyc-solutions/problems-challenges-traditional-kyc-systems/> accessed 23 April 2023.

This linkage with Aadhaar was facilitated through India Stack, *"a set of APIs[14] that allows governments, businesses, startups and developers to utilise a unique digital Infrastructure to solve India's hard problems towards presence-less, paperless, and cashless service delivery."*[15] This Stack, the first of several other emerging stacks in India leveraging digital identity has a cashless, paperless, presence less and consent layer, intended to enable APIs for Aadhaar Authentication and eKYC developed by the UIDAI, eSign meant for digital signature developed by the Controller for Certifying Authorities, UPI developed by National Payments Corporation of India, among others. The Open API policy forms the basis of both India Stack and National Health Stack services. Open APIs, in their existing form in government applications, allow interoperability between different e-Governance applications. Despite these initiatives, the source code of such applications has not been made available under the open-source license, thus not making it possible to be tested and audited openly. Parts of this centralised digital infrastructure remain proprietary. It is still prescriptive of what kind of solutions can be built upon it. Hence, while the India Stack and the National Health Stack are built on open APIs, they offer limited opportunity opportunities for other stakeholders to build different kinds of services. Further, the infrastructural requirements assumed currently for such an ecosystem to function effectively, do not match the infrastructural availability on the ground. Thus, services such as UPI, e-Sign, and e-KYC would still be inaccessible to a large section of the population, as they require access to a mobile phone and network connectivity. For the purposes of FinTech lending, there are two APIs within India Stack that were particularly relevant; the first is e-KYC[16] which embraces India Stack's paperless goal, by verifying the identity and address of a person through Aadhaar authentication. The second is e-Sign, which *"allows an Aadhaar holder to electronically sign a form/document anytime, anywhere, and on any device legally in India."*[17]

---

[14] API stands for Application Programming Interface, which is essentially a set of clearly defined methods of communication between various software components. For further reading, please see 'What is an API? In English, Please' (*freeCodeCamp*, 19 December 2019) <https://medium.freecodecamp.org/what-is-an-api-in-english-please-b880a3214a82> accessed 23 April 2023.

[15] 'About India Stack', <https://indiastack.org/about/> accessed 23 April 2023, 'Whats is the India Stack?; All You Need to Know' *Times Now* (13 February 2023) <https://www.times-nownews.com/technology-science/whats-is-the-india-stack-all-you-need-to-know-article-97860756#:~:text=According%20to%20the%20official%20website,%2C%20and%20cashless%20service%20delivery.%22> accessed 23 April 2023.

[16] 'India Stack's explanation on E-KYC' <https://indiastack.org/ekyc/> accessed 23 April 2023.

[17] 'India Stack's Explanation on E-SIGN'<https://indiastack.org/esign/> accessed 23 April 2023.

The use of Aadhaar-based authentication in the FinTech sector brought down the cost of onboarding substantially, making smaller loans economically viable for lenders, and opening up the market to "*previously underserved communities*."[18] The emergence of India Stack allowed apps to authenticate new customers via Aadhaar's eKYC, an online authentication mechanism linked to people's unique Aadhaar IDs, and also leverage UPI a real-time money transfer protocol.[19] Together, they dramatically reduced the costs of both onboarding customers and transfer of funds for online businesses.

Even so, the Aadhaar system has been a controversial topic of public debate since its inception for multiple reasons. It has worrying implications for the enjoyment of fundamental rights enshrined in the Indian Constitution, particularly the right to privacy.[20] Aadhaar has also come under focus for having major security flaws,[21] with multiple leaks being revealed over the years.[22] There are also privacy implications of India Stack, as services such as eKYC and UPI collect sensitive data of residents during transactions. The financial data allows more power to banks and other financial institutions, as it can be used for creating credit profiles of residents.[23] The ability of the project to meet its goals of unique identification through biometric authentication has also been strongly critiqued over the years following costly errors sometimes leading to loss of life, insecure software and multiple hacks.[24]

---

[18] PP Thimayya, 'India Stack to Serve the Underserved', *The Financial Express* (August 2017) <https://www.financialexpress.com/industry/technology/india-stack-to-serve-the-underserved/821926/> accessed 23 April 2023.

[19] Rohin Dharmakumar, 'Aadhaar and the Gradual Collapse of India Stack Live by Aadhaar, Die by Aadhaar', <https://the-ken.com/story/aadhaar-and-the-gradual-collapse-of-india-stack/>.

[20] Amber Sinha and Pranesh Prakash, 'Privacy Concerns Overshadow Monetary Benefits of Aadhaar Scheme' *The Hindustan Times* (New Delhi, 12 March 2017) <https://www.hindustantimes.com/india/privacy-concerns-overshadow-monetary-benefits-of-aadhaar-scheme/story-E3o0HRwc6XOdlgjqgmmyAM.html> accessed 23 April 2023.

[21] Usha Ramanathan, 'All is not well with Aadhaar' *The Indian Express* (7 January 2018) <https://indianexpress.com/article/opinion/columns/all-is-not-well-with-aadhaar-leak-aadhaar-details-5013305/> accessed 23 April 2023.

[22] Amber Sinha and Srinivas Kodali, 'Information Security Practices of Aadhaar (or lack thereof): A Documentation of Public Availability of Aadhaar Numbers with Sensitive Personal Financial Information' (*The Centre for Internet and Society*, 16 May 2017) <http://cis-india.org/internet-governance/information-security-practices-of-aadhaar-or-lack-thereof-a-documentation-of-public-availability-of-aadhaar-numbers-with-sensitive-personal-financial-information-1> accessed 23 April 2023.

[23] Shashidhar KJ, 'Privacy International Raises Concerns over IndiaStack & UPI for Establishing Financial Identity' (*Medianama*, 4 December 2017) <https://www.medianama.com/2017/12/223-privacy-international-upi-indiastack/> accessed 23 April 2023.

[24] Reetika Khera, 'Aadhaar Failures: A Tragedy of Errors', Economic & Political Weekly (2019) 54 (16) <https://www.epw.in/engage/article/aadhaar-failures-food-services-welfare> accessed 23 April 2023.

In the case of FinTech companies using Aadhaar for onboarding customers - each case of authentication *and authorisation creates a digital trail,* providing the government, and (in the absence of adequate safeguards in the law) potentially private parties with access to granular information about intimate details of individual's lives. While the central repository of the Aadhaar ecosystem maintained by the UIDAI may be more secure, the project has also led to the creation of an ecosystem built around the digital identity framework where other public and private actors also interact with the identity program leading to exponentially increased generation of data. The API-based system means that various actors are involved in building services on top of the Aadhaar identity layer. Further, the seeding of other public databases with Aadhaar numbers also meant that personal and sensitive data held by other government operations was now integrated with Aadhaar data. Even if we assume that the CIDR, the central repository which houses the enrolment data including biometrics is secure, the nodal points which engage with Aadhaar data, and often involve collection, storage, access to and processing of Aadhaar numbers, biometrics and connected profiling data often lack similar technological or process protections. These include cybersecurity protections, strict processes such as access control and severe penal provisions.

Particularly in the context of e-KYC, the privacy implications of Aadhaar authentication became a cause of grave concern following the passage of the Aadhaar Act in 2016.[25] Before 2016, the CIDR was only meant to provide a "yes" or "no" answer for the purpose of authentication. This was also explicitly provided for in the National Identification Authority of India Bill 2010 (NIAI) which contemplated only these two responses from the CIDR, *"The Authority shall respond to an authentication query with a positive or negative response or with any other appropriate response excluding any demographic information and biometric information."*[26] While this Bill did not become law, the Aadhaar Act that was passed in 2016 removes the safeguards contemplated in the NIAI Bill 2010. Under the Aadhaar Act, the CIDR is now permitted to respond with, *"a positive, negative or any other appropriate response sharing such identity information excluding any core biometric information."*[27] What is particularly worrying in light of this change is that the term "appropriate response" is not defined, leaving it susceptible to wide interpretation, which could *prima facie* include

---

[25]   The Aadhaar (Targeted Delivery of Financial and Other Subsidies, Benefits, and Services) Act, 2016, s 57.

[26]   The National Identification Authority of India Bill ('NIAI') 2010, s 5(2).

[27]   The Aadhaar (Targeted Delivery of Financial and Other Subsidies, Benefits, and Services) Act 2016, s 8(4).

demographic information. Therefore, from a pure authentication system which would ensure a degree of data minimisation, the possibility that requesting entities could access more identity information without user consent is built into this process. This added provision also belies the claim that Aadhaar is intended only for correct authentication, and demonstrates that there may be a clear intent for mission creep to use the authenticating system for greater sharing of personal data.

B. Machine learning based lending

One of the promises that FinTech lending brings to the fore is that of financial inclusion. The fact that individuals who were earlier invisible to traditional financial services and formal credit systems, are now potential customers, can be owed to the fact that FinTech companies access data about individuals that did not traditionally factor into credit decisions.[28] The ability to factor in non-traditional types of data, and look at 20,000 - 30,000 data points[29] that signal various aspects of a person's life for the purpose of assessing creditworthiness brings the promise of banking to those who previously thought they were ineligible. This is because of increasing reliance on machine learning systems that improve the performance of a task over time, at speeds and scales that are far beyond the reach of humans. To glean intimate details about a person's life from their behaviour online, and factor in these data points into building a cohesive map of an individual's life is essentially what ML systems offer in the FinTech lending sector.[30] ML-based lending, thus introduces the promise of efficiency at scale in assessing the credit-worthiness of potential customers.

The ability of ML systems to learn from examples, make inferences and spot patterns at great speeds and enormous scale contribute to the excitement surrounding the use of these systems in the financial services sector.[31] The use of ML systems for making decisions about credit, for example,

---

[28]   M.A. Bruckner, 'The Promise and Perils of Algorithmic Lenders' Use of Big Data' Chicago-Kent Law Review (2018) 93(1) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3137259> accessed 23 April 2023.

[29]   Raktim Nag, 'How Matrix Backed FinTech Startup Finomenais Disrupting the $8 Bn Youth Loan Market'(*Inc 42,* 10 June 2016) <https://inc42.com/startups/finomena/> accessed 23 April 2023.

[30]   Dirk A. Zetzsche and others, 'From FinTech to TechFin: The Regulatory Challenges of Data-Driven Finance' (2017) University of Hong Kong Faculty of Law Research Paper No. 2017/007 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2959925> accessed 23 April 2023.

[31]   Peter Martey Addo, Dominique Guegan, and Bertrand Hassani, 'Credit Risk Analysis using Machine and Deep Learning Models' (2018) University Ca' Foscari of Venice, Dept. of Economics Research Paper Series No. 08/WP/2018 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3155047> accessed 23 April 2023.

significantly changes the manner in which traditional lending takes place.[32] FinTech startups can now use new sources of data, such as social media data, or call data, to make decisions about the credit-worthiness of individuals.[33] These non-traditional types of data are often termed "alternative data". Some FinTech companies factor in up to 22,000 data points to assess credit-worthiness of individuals.[34] The impact of such technology on the overall landscape of financial services, particularly on financial inclusion is well understood.[35] However, the implications of these systems on privacy, identity, and inclusion are less thoroughly considered.

While thinking through the implications of ML systems, it is essential to understand the process through which these systems are built and deployed. In an academic paper published in 2018, Marda offers a framework for this by dividing the ML process into three distinct steps: Data, Model, and Application.[36] ML algorithms are trained on datasets often referred to as *"training data"*. For the purposes of FinTech lending, this could be datasets that contain information about people's behaviour online, their spending patterns, their living conditions, geolocation, and so on. As mentioned above, some FinTech companies in India have publicly acknowledged that the number of data points is often around 20,000.[37]

III. Data

ML-enabled credit scoring works by collecting, identifying and analysing data that can be used as proxies for information that helps answer the three key questions in any credit scoring model— a) identity, b) ability to replay and c) willingness to repay. With the advent of Big Data and greater digitization

---

[32] Matthew A. Bruckner, 'Regulating FinTech Lending'(2018) 37(6) 1 Banking & Financial Services Policy Report <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3207365> accessed 23 April 2023.

[33] Vivina Vishwanathan, 'SMS, Social Media may Reveal Credit Strength' (*Livemint*, 17 November 2015) <https://www.livemint.com/Money/9LdV0ttbYT2BgVFbLwN6UM/SMS-social-media-may-reveal-credit-strength.html> accessed 23 April 2023.

[34] Aparajita Choudhury, 'How Finomena is Making it Possible for Borrowers without Credit Scores to get a Loan', (*YourStory*, 18 February 2017) <https://yourstory.com/2017/02/finomena-2/> accessed 23 April 2023.

[35] Shekhar Lele, 'Fintech 2.0: A New Era of Financial Inclusion' (*PwC*, November 2018) <https://www.pwc.in/consulting/financial-services/fintech/fintech-insights/fintech-2-0-a-new-era-of-financial-inclusion.html> accessed 23 April 2023.

[36] Vidushi Marda, 'Artificial Intelligence Policy in India: A Framework for Engaging the Limits of Data-Driven Decision-Making' [2018] Philosophical Transactions A: Mathematical, Physical and Engineering Sciences <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3240384> accessed 23 April 2023.

[37] Nag (n 29).

and datafication of information, new data sources such as telecom data, utilities data, retailers and wholesale data and government data, are available. Examples of telecom data include prepaid data and recharge patterns that are said to provide insights about a person's cash flows. The daily call patterns and location data can indicate whether a person is working a steady job or not. One of the key sources of proxy data about income and spending is the texts about payments, and the credit and debit texts received on the consumer's mobile phone. Payment of bills, purchases made, regular remittances and made or received are all deemed very useful in predicting a consumer's ability and intent to repay.

The digitisation of records and the use of digital payment mechanisms to pay utilities bills make this data available for analysis. This data not only shows the consumption patterns of an individual but also how timely the person is in making payments. The payments cycles for utilities bills are usually periodic, like monthly repayment cycles and therefore considered highly indicative of how the person handles their monthly financial obligations. In India, electricity bills, which indicates the usage of household appliances, are widely seen as good indicators of socio-economic status and income.[38] Retailers' data can be used to evaluate the individual's expendable income, their family structure, other relevant characteristics, for instance, purchase of certain goods can suggest health consciousness while others may indicate risk taking abilities.

The metadata collected by the mobile apps used by small lending firms are analysed to derive insights about the consumer. The mobile apps typically seek various permissions to access other data on the person's mobile phone and their logon identities like Facebook and Google. Further, psychometric analysis of the manner in which the consumer fills the online form on the app, such as time taken on each question, the number of times an answer was changed etc. are also seen as indicative of the individual's character.[39]

IV. Model

As algorithms train, what emerges from the training process is called a *"model"* which is a decision matrix that can then be refined and tested till

---

[38] Shivam Shankar Singh, *How to Win an Indian Election: What Political Parties Don't Want You to Know* (Penguin Ebury Press 2019).

[39] Amber Sinha, 'Big Data in Credit Scoring', in Elonnai Hickok, Sumandro Chattapadhyay and Sunil Abraham (eds), *Big Data in Governance in India: Case Studies* (The Centre for Internet and Society 2017) <https://cis-india.org/internet-governance/files/big-data-compilation.pdf> accessed 23 April 2023.

it is considered appropriate for deployment. As models continue to be built and trained, they are deployed when they get comfortably close to a definition of "success" as laid out by the engineers who build these systems. Once this is achieved, models can be deployed for the purposes of credit scoring, underwriting, etc. This means that the definition of success, the choice of data used to train algorithms, and the criteria used to assess the performance and appropriateness of machine learning models are all extremely subjective, human decisions. This stands in strong contrast to the generally held belief that algorithmic models are all-knowing, neutral and objective.

Traditionally, credit scoring algorithms would consider set categories of data such as an individual's payment history, debt-to-credit ratio, length of credit history, new credit, and types of credit in use.[40] Machine learning algorithms as envisioned by the FinTech sector use thousands of alternate data points such as the number of contacts in one's phone, call logs, and social media behavior to discern an individual's creditworthiness.[41] The first implication of this type of model is that it is not always possible to explain why a certain decision was made, as models that use complex techniques like neural networks are inscrutable even to those individuals who build them. Given the vast amount of data analyzed and complex structures within neural nets it may not even be possible for lenders to understand why certain loan applications are approved while others are rejected. Second, creditworthiness is not easy to predict, particularly given that historical data on access to credit, payment and default is imbued with a number of societal realities along the axes of gender, class, caste, religion, and so on — complexities that datasets do not reflect. For instance, if a model is trained with data only about men receiving and repaying loans, and does not 'learn' from any examples of women being good credit prospects, this could risk women's access to credit in the future.[42]

With the introduction of new forms of data, the richness of data may theoretically increase the predictive power of the algorithm. However, narratives on greater accuracy presume both the suitability of input data towards

---

[40] See National Consumer Law Center, *Fair Credit Reporting* § 16.4.5.2, at 720 (9th edn 2017).

[41] Pierre Biscaye and others, 'Review of Digital Credit Products in India, Kenya, Nigeria, Tanzania and Uganda' (2017) EPAR Technical Report #351a <https://epar.evans.uw.edu/sites/default/files/EPAR_UW_351a_Review%20of%20Digital%20Credit%20Products_4.12.17_0.pdf> accessed 23 April 2023.

[42] In a non-lending context, Amazon's hiring algorithm made a similar mistake; See Jeffrey Dastin, 'Amazon Scraps Secret AI Recruiting Tool that Showed Bias against Women' (*Reuters*, 11 October 2018) <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G> accessed 23 April 2023.

the desired output, as well as faith that past attributes or activities that are used as training data do not lead to unintended outcomes. The use of alternative data and assumptions about proxy factors that influence ability and willingness to pay are both largely untested. Therefore, there is a risk of creating a financial market which is dependent on unproven assumptions.

V. Application

The overarching narratives around the use of machine learning in the FinTech sector are that of efficiency, and providing credit to those who were not included in traditional financial systems. Individuals with thin credit files and limited interaction with financial services were stuck in a vicious circle that denied them credit, but with the use of alternate data, this can change. This narrative also promises quicker access to credit due to the sheer speed and agility of ML systems.[43] However, the limitations of these systems are significant in the context of FinTech lending and require thoughtful deliberation.

First, ML systems that are trained for the purpose of financial services need to carefully consider the data used to train systems. Financial disparity in India is large, and thus, the choice of datasets has an impact on how these systems will function. An ML system trained on the financial behavior of predominantly affluent people, for example, will systematically underperform or exclude less affluent people because of embedded assumptions about the "ideal" case in datasets. For instance, affluent people may, on average, have a significant portion of money being transferred to equity and debt investments, which can in turn become a proxy for the "ideal" individual. This is not a luxury that individuals from less affluent sections of society necessarily have, particularly not in a country like India, thus making them immediately at odds with the "ideal" loan applicant. Similarly, communities that have been systematically excluded by social and political norms will have distinct financial footprints and behaviors. Thus, an uncritical adoption of data can lead to a situation where people continue to be discriminated against and excluded simply because historical bias is being encoded in formal and opaque ways into ML systems.

---

[43] Arjuna Costa, Anamitra Deb, and Michael Kubzansky, Big Data, Small Credit: The Digital Revolution and its Impact on Emerging Market Consumers (2015) 10 (3-4) 'Innovations: Technology, Governance Globalization' 49 <https://ideas.repec.org/a/tpr/inntgg/v10y2015i3-4p49-80.html> accessed 23 April 2023.

A preliminary survey of the emerging companies in the Fintech sector in India done in 2017 and the profiles of their management teams show a preponderance of those with 44 technology and sales background and a lack of individuals trained in banking and finance.[44] This suggests an over-reliance on data and technology, and a tendency to ignore other kinds of expertise which have been integral to the credit scoring industry. This is reflective of the narrative that data is exhaustive and comprehensive enough to provide inferences that negate the need for domain expertise, theoretical models and interpretivism. However, this assumption has been greatly critiqued and various authors have pointed out the perils of the over-reliance on data.[45] However, this ignores the need for professionals with prior domain knowledge who can critically look at the predictions or inferences made by machine learning algorithms.[46]

Second, ML systems today often lack the Indian context: A classic credit underwriting ML system is built using practices imported from developed economies, which impacts their efficacy and accuracy in the Indian context. For example, people's geolocation and their call data records are thought to reveal a lot about their personalities and lifestyle. However, this assumption is lost in the context of loan applicants who are women from traditional families in some parts of India —a cellphone is not a personal possession, but rather a household one, often in the name of the head of the family who is invariably a man. This means that perfectly good candidates who deviate from the norm of what is considered "normal" behaviour in the West run the risk of being systematically excluded by these systems. Contextual development of models is key, failing which these systems could end up excluding vulnerable communities as the norm.

Finally, ML systems have profound implications for **privacy and autonomy**. From inferring intimate details about an individual's life, to potentially enabling surveillance, even well-intentioned ML systems can be detrimental to privacy. Further, the volumes at which these systems are trained mean that multiple correlations can emerge, some of which may pertain to sensitive

---

[44] Sinha (n 39).

[45] S. Leonelli, 'What Difference does Quantity Make? On the Epistemology of Big Data in Biology' (2014) 1(1) Big Data & Society <https://journals.sagepub.com/doi/epub/10.1177/2053951714534395> accessed 23 April 2023. ; Fulvio Mazzocchi, 'Could Big Data be the End of Theory in Science? A few Remarks on the Epistemology of Data-Driven Science (2015) 16(10) EMBO Reports1250 <https://doi.org/10.15252/embr.201541001> accessed 23 April 2023.

[46] Mireille Hildebrandt, 'Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning' (2019) 20(1)83 <https://doi.org/10.1515/til-2019-0004> accessed 23 April 2023.

attributes. Even if the correlation is only very slight, this is enough to build systems that factor in sensitive attributes. It is also unclear whether the use of ML has helped with access to credit in a sustainable way, and if financial inclusion is meaningfully achieved at all. The models, datasets, and applications that are currently in play are not subject to audits, with no transparency or accountability mechanisms.

The FinTech sector has grown substantially in the last few years because of these two factors, i.e. the option of Aadhaar-based authentication, and the growth in popularity of machine learning techniques and applications. As this Section demonstrates, however, the adoption of these technical 'solutions' cannot and should not be treated as straightforward or simplistic - particularly in a country like India, where the layers of complexity and disparity merit a close, deliberate and careful approach to critical services such as access to credit.

VI. RECENT REGULATORY DEVELOPMENTS

Having discussed the promises and limitations of these two underlying technologies in the FinTech sector, we will now turn to a brief analysis of a few developments from the last few years that had an impact on companies' ability to use these technologies - the Aadhaar judgment and the Personal Data Protection Bill.

A. Aadhaar judgment

In September 2018, the Supreme Court of India upheld the constitutional validity of Aadhaar with a 4:1 majority, following the second longest hearing in the Supreme Court's history.[47] While the judgment covers a range of important and intricate issues from proportionality to surveillance, for the purposes of this paper, we will discuss the extent to which private parties' use of Aadhaar was curtailed, what questions remain, and what the status quo is.

In discussing the use of Aadhaar by private companies, Section 57 of the Aadhaar Act came into focus and was found to be unconstitutional by all three opinions that made up the judgment. This section allowed for the use

---

[47] Moneylife Digital Team, 'Historic Aadhaar Hearing, Second-longest in SC history, Concludes' (*Money life*,10 May 2018) <https://www.moneylife.in/article/historic-aadhaar-hearing-second-longest-in-sc-history-concludes/53992.html> accessed 23 April 2023.

of Aadhaar in establishing the identity of a person for **any** purpose, by a state or a body corporate or person.

While discussing this section of the Act, the majority found[48] that it was susceptible to misuse as:

"*(a) It can be used for establishing the identity of an individual 'for any purpose'. We read down this provision to mean that such a purpose has to be backed by law. Further, whenever any such "law" is made, it would be subject to judicial scrutiny.*

(b) Such purpose is not limited pursuant to any law alone but can be done pursuant to 'any contract to this effect' as well. This is clearly impermissible as a contractual provision is not backed by a law and, therefore, first requirement of proportionality test is not met.

(c) Apart from authorising the State, even 'any body corporate or person' is authorised to avail authentication services which can be on the basis of purported agreement between an individual and such body corporate or person. Even if we presume that legislature did not intend so, the impact of the aforesaid features would be to enable commercial exploitation of an individual biometric and demographic information by the private entities. Thus, this part of the provision which enables body corporate and individuals also to seek authentication, that too on the basis of a contract between the individual and such body corporate or person, would impinge upon the right to privacy of such individuals. This part of the section, thus, is declared unconstitutional."

While discussing Section 57, Justice Bhushan found, "*When any law permits user of Aadhaar, its validity is to be tested on the anvil of threefold test as laid down in Puttaswamy case, but permitting use of Aadhaar on any contract to this effect, is clearly in violation of Right of Privacy. A contract entered between two parties, even if one party is a State, cannot be said to be a law. We thus, are of the view that Section 57 in so far as it permits use of Aadhaar on "any contract to this effect" is clearly unconstitutional and deserves to be struck down.*"[49]

Finally, the dissenting opinion from Justice Chandrachud found "*Section 57 indicates that the legislature has travelled far beyond its stated object of ensuring targeted delivery of social welfare benefits. Allowing the Aadhaar*

---

[48]   K.S. Puttaswamy v Union of India (2017) 10 SCC 1. A.K. Sikri, J. p 561.

[49]   K.S. Puttaswamy v Union of India (2017) 10 SCC 1. Ashok Bhushan, J. Para 282, p 264.

platform for use by private entities overreaches the purpose of enacting the law. It leaves bare the commercial exploitation of citizens data even in purported exercise of contractual clauses. This will result in a violation of privacy and profiling of citizens." He further stated, *"Section 57 does not pass constitutional muster. It is manifestly arbitrary, suffers from overbreadth and violates Article 14."*[50]

Following the judgment, FinTech firms had to grapple with alternatives to e-KYC that can offer similar ease of execution and .cost-effectiveness. At the time of the judgment being pronounced, there was a sense of doom within the FinTech industry.[51] Following this, the UIDAI offered two alternatives to continue using Aadhaar without sharing biometric information or the Aadhaar number - by either using a QR code[52] or a digitally signed XML file.[53] A few months down the line, it was clear that some types of lenders are hit more than others.[54] Lenders focusing on short-term, small ticket loans of less than one lakh, simply have not found economically viable options as traditional KYC costs are too high, and in the meanwhile are moving towards video-KYC,[55] and other methods through dialogue with regulators. On the other hand, lenders who are more diversified in the market seem to be embracing alternatives, such as more traditional banking KYC methods which rely on paper documents such as PAN and Driver's License. This form of authentication usually employs the Original Seen and Verified ("OSV") method where the original copy of the document should be seen and verified by the case officer. UIDAI also introduced its offline verification tools like XML databases and QR code-based solutions.

---

[50] K.S. Puttaswamy v Union of India (2017) 10 SCC 1. Dr Dhananjaya Y Chandrachud, J. Para 245, p 338.

[51] Vanita D'Souza, 'Here is Why the Aadhaar Verdict Left Fintech Companies in Ripples' (*Entrepreneur*, 23 December 2018) <https://www.entrepreneur.com/article/325288> accessed 23 April 2023.

[52] Mayur Shetty, 'Banks may Use Aadhaar QR Code for Paperless KYC' *The Times of India* (New Delhi, 26 October 2018) <https://timesofindia.indiatimes.com/business/india-business/banks-may-use-aadhaar-qr-code-for-paperless-kyc/articleshow/66370303.cms> accessed 23 April 2023.

[53] UIDAI, 'Offline Aadhaar Data Verification Service' (*UIDAI*, 23 August 2018) <https://uidai.gov.in/images/Offline-Aadhaar-Data-Verification-Service_v1-23082018.pdf> accessed 23 April 2023.

[54] Pratik Bhakta, 'India's FinTech Companies Struggle for an Alternative to Aadhaar' *The Economic Times*(21 December 2018) <https://economictimes.indiatimes.com/small-biz/startups/features/indias-fintech-companies-struggle-for-an-alternative-to-aadhaar-articleshow/67186586.cms> accessed 23 April 2023.

[55] Shreya Ganguli, 'RBI Mulls Live Video Authentication as Aadhaar eKYC Alternative' (*Inc42*, 10 December 2018) <https://inc42.com/buzz/rbi-mulls-live-video-authentication-as-aadhaar-ekyc-alternative/> accessed 23 April 2023.

While the judgment clearly finds Section 57 to be unconstitutional, there has been some speculation on the extent to which private players can use Aadhaar for e-KYC going forward. Shortly after the verdict was pronounced, Finance Minister Arun Jaitley stated that if the use of Aadhaar for private players *"is backed by a law, it is not unconstitutional."*[56] There have been legislative efforts to revive Aadhaar-based e-KYC for private parties through the Aadhaar and Other Laws (Amendment) Bill 2018.[57] This contemplates making furnishing Aadhaar "voluntary", and proposes amendments to the Prevention of Money Laundering Act and the Telecom Act, by allowing Aadhaar access to banks and telecom operators. Elsewhere, FinTech companies sought clarification on whether the use of e-KYC by them would be permitted if it was done on a voluntary basis.[58] There have been two views about the extent of the application of reading Section 57 down by the court. The first view posited that this meant that "private actors were not permitted to use the Aadhaar infrastructure even as requesting entities, even under a voluntary contract."[59] On the other hand, the second view argues that the wide definition of the term 'requesting entity' in the Aadhaar Act and the UIDAI's power to authenticate the request of any requesting entity also includes private sector parties.[60]

In July 2019, the Rajya Sabha passed the Aadhaar (and other laws) Amendment Bill.[61] In line with the Aadhaar judgment, Section 57 was omitted, however Section 4(4), Aadhaar Act was introduced to permit "an entity" to perform authentication, as long as (i) it was compliant with certain specified standards of privacy and security (which are yet to be specified) and (ii) it was permitted to offer authentication services by law or it was seeking

---

[56] Karan Dhar, 'Arun Jaitley Hints at New Law after Supreme Court Bars Private Companies from Using Aadhaar Data' (*Business Today*, 26 September 2018) <https://www.businesstoday.in/current/economy-politics/arun-jaitley-aadhaar-supreme-court-private-companies-banks-law/story/282886.html> accessed 23 April 2023.

[57] The Aadhaar and other Laws (Amendment) Bill 2019 <https://prsindia.org/files/bills_acts/bills_parliament/2019/Aadhaar%20and%20Other%20Laws%20(Amendment)%20Bill,%202019.pdf> accessed 23 April 2023.

[58] Yuthika Bhargava, 'FinTech Companies Seek Clarity on Using Aadhaar for e-KYC' *The Hindu* (New Delhi, 14 December 2018) <https://www.thehindu.com/business/fintech-companies-seek-clarity-on-using-aadhaar-for-ekyc/article25746312.ece> accessed 23 April 2023.

[59] Vrinda Bhandari, 'Governing ID: India's Unique Identity Programme' (*Digital Identities and Uses*, 6 February 2023) <https://digitalid.design/evaluation-framework-case-studies/india.html>accessed 23 April 2023.

[60] The Aadhaar (Targeted Delivery of Financial and Other Subsidies, Benefits and Services) Act 2016, s 8(1).

[61] Aadhaar (and other Laws) Amendment Act 2019 (*PRS Legislative Research*) <https://www.prsindia.org/billtrack/aadhaar-and-other-laws-amendment-bill-2019> accessed 23 April 2023.

authentication for certain prescribed purposes. Through this legislation, the private sector use of Aadhaar was effectively restored. The Telegraph Act and the PMLA Act were also amended to allow various private entities to use Aadhaar for authentication.

In light of the Supreme Court's verdict discussed above, and consequent efforts to revive Aadhaar authentication for private companies, the use of Aadhaar for e-KYC going forward, in our view, will be impermissible even through a new law, given that the crux of such access involves: 1) commercial exploitation of an individual's sensitive personal information, including biometric and demographic information; and 2) through (voluntary)[62] contracts - the very basis on which the court struck down Section 57 in the first place. The legal rationale behind striking down the use of Aadhaar under Section 57 relies on the age-old dictum that what is prohibited by law, cannot be facilitated by way of contract. Section 57 played the role of carving out an entire ecosystem of contractual transactions, outside the purview of protections and governance in the Act. It is this carve out that the Supreme Court struck down, and has been reinstated contrary to the spirit of the Aadhaar judgment through the 2018 rules.

VII. Personal Data Protection Bill

The first version of the Personal Data Protection Bill[63] **was published in July 2018, along with the final** report[64] of the Justice Srikrishna Committee on Data Protection. Over the last four years, two subsequent versions of the bill, one from MeitY[65] and another from the Joint Parliamentary Committee haveemerged.[66] In each of these draft legislations, informed consent remains the primary ground for the processing of personal data. Although it must be noted that the scope of non-consensual grounds has only increased in each subsequent draft.

---

[62] Prasanna S, 'Section 57: Why Aadhaar can't be Used as Authentication by Private Companies' (*Medianama*, 27 September 2018) <https://www.medianama.com/2018/09/223-section-57-why-aadhaar-cant-be-used-as-authentication-by-private-companies/> accessed 23 April 2023.

[63] The Personal Data Protection Bill 2018 <http://meity.gov.in/writereaddata/files/Personal_Data_Protection_Bill,2018.pdf> accessed 23 April 2023.

[64] BN. Srikrishna and others, 'A Free and Fair Digital Economy Protecting Privacy, Empowering Indians: Committee of Experts under the Chairmanship of Justice B.N. Srikrishna' (*Ministry of Electronics and Information Technology*, 27 July 2018) <http://meity.gov.in/writereaddata/files/Data_Protection_Committee_Report.pdf> accessed 23 April 2023.

[65] Personal Data Protection Bill 2019 <http://164.100.47.4/BillsTexts/LSBillTexts/Asintroduced/373_2019_LS_Eng.pdf> accessed 23 April 2023.

[66] Report of Joint Committee on the Personal Data Protection Bill 2019.

For FinTech lending companies, the notion of informed consent is one that needs to be studied more closely. Most lenders obtain explicit consent from customers, by obtaining signatures and multiple consent forms as part of the onboarding process. The extent to which this consent is informed, free and specific is limited. For consent to be informed, when given in response to written declaration which also concerns other matters, requires the request for consent shall be presented in a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language. For consent to be free, we need to consider whether the performance of a contract or provisions of service is conditional on consent being provided to a non-negotiable, one-sided contract. This is particularly relevant in the context of alternative data given that lenders who factor in multiple data points from all aspects of an individual's life are, it can be argued, essentially carrying out a business model that is at odds with the purpose of data minimization and collection limitation. Another aspect to consider is the limitations on the storage of personal data, with the law contemplating that data fiduciaries *"shall retain personal data only as long as may be reasonably necessary to satisfy the purpose for which it is processed."*

While there are undoubtedly significant improvements made to the data protection landscape through the various versions of the Personal Data Protection Bills, a pessimistic reading of the draft legislations leads to the conclusion that it may not go too far in protecting consumers' data in the context of FinTech lending for two reasons:

First, while there is a requirement for informed and explicit consent, the latter is slowly becoming a surrogate for the former, particularly in the absence of existing mechanisms that explain how to operationalize informed consent in the context of FinTech lending. Second, the Bill does not talk about privacy considerations at the level of machine learning models, unlike the rights on automated processing and explanation provided in the EU's GDPR. This effectively means that models can continue to be opaque even once the Bill comes into force, and be built and deployed in a manner that is detrimental to the right to privacy of individuals.

Practices such as checking credit scores during background verification for employment, health insurance etc. have been criticized for a long time. However, big data-enabled credit scoring provides a far more granular profile involving different behavioral aspects of a person and the big data ecosystem provides more opportunities for credit data to be used for non-credit purposes. In light of the lack of regulation in the Fintech sector, there is a

risk of such practices emerging as a business model to generate additional revenue for the companies.

VIII. DIGITAL LENDING REGULATIONS

On September 2, 2022, the Reserve Bank of India (RBI) released a set of guidelines to regulate digital lending in India.[67] This followed a framework released by RBI in August 2022.[68] The regulations introduce some regulatory restrictions on digital lending apps. First, it introduces privacy protections for data collection carried by service providers. It requires that processing should be need-based with clear audit trails, and should be only done with the prior explicit consent of the borrower. In order to address, blanket app permission taken by such services, it imposes restrictions on access to mobile phone resources such as files and media, contact lists, call logs, and telephony functions. Further prescriptive provisions require that one-time access can be taken for the camera, microphone, location or any other facility necessary for the purpose of onboarding or KYC requirements only with the explicit consent of the borrower. Other obligations include the need for a privacy policy, data localisation, data security, transparency around data storage etc.

The second set of rules relevant for our discussion here is the obligation to ensure that the algorithm used for underwriting is based on extensive, accurate and diverse data to rule out any prejudices. RBI also imposes auditability requirements for the algorithm up to minimum underwriting standards and potential discrimination factors used in determining credit availability and pricing. In the same vein, the regulations encourage ethical AI which focuses on protecting customer interest and promotes transparency, inclusion, impartiality, responsibility, reliability, security and privacy. These are early attempts towards regulating predatory practices in the lending industry and will require significant fine-tuning and evolution. The first impressions of the industry have been largely negative towards the rules, with concerns around the prescriptive nature of the provisions.[69] One technology lawyer

---

[67] Reserve Bank of India, 'Guidelines on Digital Lending' (*RBI*, 2 September 2022) <https://www.rbi.org.in/Scripts/NotificationUser.aspx?Id=12382&Mode=0> accessed 23 April 2023.

[68] Reserve Bank of India, 'Recommendations of the Working Group on Digital Lending - Implementation' (*Reserve Bank of India*, 10 August 2022) <https://www.rbi.org.in/Scripts/BS_PressReleaseDisplay.aspx?prid=54187> accessed 23 April 2023.

[69] Reuters, 'India's Digital Lending Rules Spark Disruption, Firms Plan Pushback' *The Economic Times* (26 August 2022) <https://economictimes.indiatimes.com/tech/technology/indias-digital-lending-rules-spark-disruption-firms-plan-pushback/articleshow/93798112.cms> accessed 23 April 2023.

that the authors spoke to indicated that the Fintech industry was likely to ramp up lobbying efforts in response to the new rules, and advocated a lighter set of laws based on first principles. It remains to be seen how these political economic factors lead to the crystallization of financial regulation of digital lending. However, it must be noted that the RBI guidelines are only a threadbare first step towards regulation of the algorithmic lending regulations with very attention paid to anti-discrimination provisions. We will look at comparative regulations below.

Beyond analyzing recent regulatory developments, it is also important to briefly touch upon the absence of legal safeguards in the context of lending. In countries like the United States, for instance, the Equal Credit Opportunity Act, 1974 (ECOA Act) prohibits discrimination on the basis of certain protected characteristics like gender, race or marital status.[70] The ECOA also protects against policies that have a disproportionate impact on protected groups (also known as protecting against disparate impact).[71] It also institutes notice requirements which compel lenders to explain why they take 'adverse action' which includes refusal to grant credit, or refusal to increase the amount of credit available to an applicant.[72] In India, the Reserve Bank of India's Guidelines on Fair Practices Code for Lenders, 2003 suggests that lenders should not discriminate on the basis of caste, sex or religion, and also requires lenders to convey in writing *the main reason/reasons which, in the opinion of the bank after due consideration, have led to rejection of the loan applications within stipulated time.*[73] However, these are merely recommendary guidelines, as recent research has found that FinTech companies in India do not readily disclose the reasons for the rejection of a loan.[74] The absence of binding regulation in India means that there are little to no safeguards in place for borrowers.

---

[70] Brian Kreiswirth and Anna-Marie Tabor,'What you need to know about the Equal Credit Opportunity Act and How it can Help you: Why it was Passed and What it is' (*Consumer Financial Protection Bureau*, 31 October 2016) <https://www.consumerfinance.gov/about-us/blog/what-you-need-know-about-equal-credit-opportunity-act-and-how-it-can-help-you-why-it-was-passed-and-what-it/> accessed 23 April 2023.

[71] Tarunima Prabhakar and Steve Weber, 'Financial Inclusion as a Fairness Criterion in Credit Risk Assessment' (*SSRN,* 25 June 2020) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3579695> accessed 23 April 2023.

[72] 'Interactive Bureau Regulations: 12 CFR Part 1002 (Regulation B)' (Consumer Financial Protection Bureau) <https://www.consumerfinance.gov/rules-policy/regulations/1002/9/> accessed 23 April 2023.

[73] Reserve Bank of India, 'Guidelines on Fair Practices Code for Lenders' (*RBI*, 5 May 2003) <https://www.rbi.org.in/Scripts/NotificationUser.aspx?Id=1172&Mode=0> accessed 23 April 2023.

[74] Tarunima Prabhakar and Steve Weber, 'Alternative Lending in a Digital Age: A Comparative Case Study in Regulation Across India and the United States' (*SSRN*, 19 May 2020) 22 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3956623> accessed 23 April 2023.

IX. How the industry coped

In order to bridge the gap between legal analysis and policy implications on one end, and practitioner perspectives on the other, this section will synthesize findings from six in-depth qualitative interviews[75] with entrepreneurs from the FinTech industry. Interviewees were founders and/or CEOs of leading FinTech lending companies in terms of market size in India that focus on easy and quick disbursal of loans, some entirely online, based on alternate data. These interviews were conducted to understand how FinTech players view policy changes in re: Aadhaar and data protection in context of their businesses, and how they have adapted to them. Interviews were semi-structure, but broadly, questions centered around *inter alia* understanding how alternate data featured in their business models, how FinTech companies view regulatory developments and whether there were certain bright lines for what definitely does or doesn't work for them, and how they coped with meaningful alternatives to Aadhaar onboarding.

The threads that emerged from interviews are discussed below:

1. **Viable alternatives to Aadhaar are possible and also feasible:** All interviewees acknowledged that the inability to use Aadhaar for onboarding customers, while inconvenient and most certainly a setback, was not fatal for most lending firms as several viable alternatives could be developed. This is for two main reasons. The first is that e-KYC was only allowed for loans up to Rs. 60,000 and some interviewees' firms only began lending at 1,00,000. For those interviewees who focus on smaller, shorter personal loans, methods of digital lending that don't require Aadhaar are currently being built and tested by the sector. Some interviewees mentioned working towards driver license or voter ID based verification, which one interviewee claimed would be "just as robust" as Aadhaar. Other lenders are moving towards video-KYC which is recognized by SEBI,[76] or an email-based KYC which works with some Non-Banking Finance Companies (NBFCs). For loans above a lakh, the KYC requirement of Original Seen and Verified (OSV) continues as it did before.

---

[75] The interviews have been completely anonymized in the interest of uniformity for this section.

[76] 'SEBI Comes Out with Revised KYC Norms for FPIs', *The Economic Times* (21 September 2018) <https://economictimes.indiatimes.com/markets/stocks/news/sebi-comes-out-with-revised-kyc-norms-for-fpis/articleshow/65902690.cms> accessed 23 April 2023.

However, after the new Aadhaar regulation circumvented the Aadhaar judgment, Aadhaar-based authentication again became the default for digital lending companies.

2. **UIDAI's offline verification tools do not inspire confidence:** The solutions offered by the UIDAI following the Supreme Court verdict, like the Offline Aadhaar XML file, or the QR code route do not seem like practical options for any of the interviewees. In case of the XML file, interviewees mentioned that it puts the onus on end customers to be digitally savvy. As one interviewee remarked, *"It may work for some, but it is not a solution for the masses"*. The QR code route has also failed to garner much excitement as lenders believe that if they are made to go to a potential lendee's house anyway, they'd much rather see a PAN card or driver's license for the purpose of KYC. Simply put, if these are the only two alternatives to eKYC, one interviewee succinctly stated, *"the economics of lending don't make sense for the small loan segment anymore"*.

3. **Overall positive response to the letter, but not necessarily the spirit of the Personal Data Protection Bill:** Most FinTech firms seem unperturbed by the standards on collection, processing, consent, and sharing introduced by the different versions of the data protection bills. One interviewee, in particular, welcomed the Bill as the *"right direction for India to move in, because the way data is handled in India today is shocking"*. He also stated that the requirements under the Bill, as far as they require specificity and security, should ideally be routine hygiene for FinTech companies. He added that another positive aspect of the bill is that it does away with 'fly-by-night operators' who collect vast quantities of data for no clear purpose. Another interviewee welcomed the fact that the bill signals the *"ecosystem is evolving to bring clarity into what can/can't happen"*. A fourth interviewee was agnostic to what the bill entails as the bill would apply to all FinTech companies equally, with no significant repercussions for competitiveness within the sector.

On the question of how requirements of consent, data minimization, purpose limitation, and collection limitation affect the sector, 5 out of 6 interviewees believed that it would change how they conducted their business. Most interviewees (save one) explained that there is explicit consent secured at the time of onboarding. One interviewee even told us that at the time of onboarding a new customer, there are approximately 40 consent forms that must be signed for the purposes of receiving credit, effectively covering all

bases on what data would be collected and processed. Another interviewee clarified that customers are free to revoke consent with ease at any given time. It is clear that these companies are focused on *explicit*consent, however, the understanding of what constitutes *informed consent*leaves much to be desired. This tension was succinctly captured by one interviewee who asked me, *"How is this new requirement of explicit consent different from a simple tick box?"*.

4. **Significantly diverging views on what constitutes alternative data:** Three of six interviewees expressed scepticism about the extent to which alternative data is used in the sector today, while two other interviewees' business model is predicated on it. However, it also appears that the definition of alternative data as understood in the industry is changing quite rapidly. One interviewee said, *"As we move away from manual underwriting, nothing is really alternative anymore. In a sense, we are using traditional data in non-traditional ways: we assess loan applications in alternative ways when you compare us to traditional financial institutions. In order to do this, we look at signals from your life and your business as a means to understand your ability and intent to pay."* This was echoed by another interviewee, who stated that *"much of what is thought of as alternative data is really mainstream data"*, and he further added, *"When people talk about alt data, they often mean traditional data through alternative means"* while referring to the use of SMS data to understand financial transactions. These interviewees look at SMS data as a proxy to official bank statements for those individuals who aren't embedded in the formal banking system. This sentiment is in sharp contrast to responses from other interviewees who use alternative data and view it as a central factor in their business model for the purpose of underwriting and lending.

5. **Significantly diverging views on the potential value of alternative data:** As mentioned above, two interviewees view alternative data as their bread and butter. The role that alternative data plays, according to one interviewee, is enabling the building of accurate prediction-based risk models and other decision engines that can inform complex decisions. Another interviewee explained that alternative data plays a crucial role in his business as the market that the company hopes to serve includes those who are not embedded in formal financial systems. On the other hand, other interviewees held a very different view - that *"social media data was a hype a few years ago, but there has been no value found from using it thus far.."* Additionally, Facebook's

move to cut off access to social media data[77] means that this is also logistically difficult to do at this time. The same interviewee added, *"The thesis for using many data points was that this could be used to include the large unbanked population of our country. But people are finding very little correlation between social media data and credit behavior. There is hard to prove incremental value, if any, of using alternatives."*Two of the six interviewees also expressed caution against the use of alternative data because they believe that losing access to alternative data is only a matter of time.

6. **Restriction on Android apps on data scraping changes very little for lenders:** At the time of Google's decision to limit third-party apps' access to user data, there was a flurry around its significant impact on lender's ability to carry out business. Less than six months after that announcement, interviewees are not worried about this shift, since Google continues to allow scraping "relevant" data for lending. One interview remarked, *"Currently, you need to justify why you need certain permissions - in this way self-regulation is making sure that data is used by the right parties in the right manner - this is both progressive and positive"*. The deficiencies of this case-by-case assessment, however, don't fix the wider issue of problematic business models that have implications for privacy. As another interviewee stated, *"This doesn't have much of an effect on how privacy is violated because some FinTech companies require contact details to call your friends and relatives at the time of collection if you are a defaulter - that will still be allowed under the justification model."* Another glaring shortcoming of this justification model was brought up by an interviewee who said, *"If you can continue using alternate methods to access traditional information… this leaves the question of other sensitive information like income tax messages being read by FinTech apps"*.

7. **Paradox of machine learning** - At the time of commencing interviews, we took the use of machine learning to be a given in this sector, but interviews indicated otherwise. One interviewee expressed scepticism around the actual use of machine learning systems in the FinTech sector in India, stating, *"From my conversation with many leaders in this space, my understanding is that there are very few use cases where ML is being used. Basic data modeling has always happened*

---

[77] Johnny Lieu, 'Facebook Cuts Off Access to User Data for 'Hundreds of Thousands' of Apps' (*Mashable India*, 31 July 2018) <https://mashable.com/article/facebook-user-data-apps/> accessed 23 April 2023.

- but no one seems to be using alternative data points to underwrite consumers. This is far from being the norm." At the same time, another interviewee explained how machine learning is a central consideration in his company's business model, and shared the three main challenges that the company faces while implementing machine learning systems. The first is that feedback cycles for machine learning models are somewhere between 9 - 12 months, which means that it takes a long time to build good credit-scoring models. Second, building models require large amounts of data, and FinTech companies in India can't build deep learning models as there isn't access to the kind of volume required for it. Finally, he mentioned that ML research and talent is funded by big tech companies that focus on certain types of problems, as a result of which *"there hasn't been an improvement in algorithms catering to the need of Indian problems and Indian consumers. There is no funding for home grown tech that takes Indian problems seriously."*

X. CONCLUSION

Through this report, we have attempted to examine the current state of FinTech lending companies in India, in the context of developments in law and policy since 2018. By offering an explanation of how Aadhaar authentication and machine learning are relevant to the sector, explaining legal developments in the context of these technologies, and informing these findings through industry interviews, we hope to have bridged the gap between legal analysis and practitioner insights.

The credit scoring industry in India needs a careful examination of rights, inclusion, appropriate safeguards and discrimination through current services. Currently, there is a lack of non-discrimination regulations that apply to the industry to safeguard against unintentional disparate impact of data-driven decision-making. There are no laws which prevent firms from collecting data on religion, caste and other sensitive attributes, which can be used toward disparate treatment. Even in other jurisdictions, there is a call for Fintech firms to be exempt from equal credit opportunity and fair credit regulations. However, regulations which prevent discriminatory practices are essential for any financial products introduced in the market.

People who lack the education, information, and other economic, cultural, and social capital that would allow them to take advantage of—and shield themselves against—the free market are most vulnerable and need greater

protection. The consequences of bad decisions are far more dire for those disadvantaged and lacking the resources—financial, psychological, social, and political—to compensate for their errors. A review of big data-enabled loan products by the National Consumer Law Centre in the US showed that they were very poor payday loan alternatives. Most of these products involved annual percentage rates three times higher than considered non-predatory. Most importantly, most products require electronic access to the applicant's bank account or some other arrangement of automatically deducting the owed amount from the borrower's account.[78]

As big data scoring uses closed and proprietary algorithm-based technologies, it is impossible to analyze them for potential discriminatory impact. There are no regulations that may be used to address discrimination on the basis of the disparate impacts of data-driven decision-making in India. The promise of Fintech lending business models to empower the unbanked and reduce timelines for approvals needs closer scrutiny. The focus of financial regulation has been on reducing financial fraud, but due to the absence of a data protection law, and non-discrimination regulations, the spillover privacy and predatory effects that are magnified by the use of machine learning algorithms are largely unregulated.

---

[78] Persis Yu, Jillian Mclaughlin, and Marina Levy, 'Big Data: A Big Disappointment for Scoring Consumer Credit Risk' (*National Consumer Law Centre*, 14 March 2014) <http://www.nclc.org/images/pdf/pr-reports/report-big-data.pdf> accessed 23 April 2023.