Bridging Data Science and Sustainable Financial Innovation

Statistics and Computing in Practice

No Author Given

Contents

Chapter 1
Foundations of Digital Finance

Joerg Osterrieder

Abstract Digital finance technologies—algorithmic trading, blockchain and decentralized finance, artificial intelligence and machine learning, digital banking and payments, and regulatory technology—are converging into a unified programmable financial infrastructure that creates both sustainability opportunities and systemic risks addressable only through data science. This chapter provides a conceptual foundation for the volume by proposing a taxonomy that organizes these five domains along shared dimensions of automation, decentralization, and data intensity. For each domain, the chapter surveys the technological foundations, key literature, and concrete linkages to sustainable financial innovation: from ESG-integrated portfolio optimization and tokenized green bonds to AI-driven climate risk disclosure and real-time supervisory analytics. A dedicated treatment of Europe's emerging Financial Data Space situates these developments within the regulatory architecture of the European Union. The chapter confronts failures and systemic concerns directly—the Terra/Luna collapse, the FTX fraud, and equity market flash crashes—analyzing how concentrated algorithmic strategies, opaque smart contracts, and fragmented oversight amplify tail risk. Sustainability receives a standalone analytical dimension, examining measurement frameworks, technological enablers, and threats posed by energy-intensive consensus mechanisms and financial exclusion. The chapter concludes by mapping open research questions onto the five thematic areas of the MSCA DIGITAL doctoral network (Grant Agreement No. 101119635), establishing the intellectual agenda that subsequent chapters develop.

Key words: Digital finance, Sustainable financial innovation, Blockchain and decentralized finance, Artificial intelligence in finance, Algorithmic trading and market microstructure, Regulatory technology, ESG measurement and sustainable finance, European Financial Data Space, Financial inclusion, Programmable financial infrastructure

JEL Classification: G14, G15, G21, G23, G28, G32, O31, O33, O36, Q54, Q56, L86, K22, D53

Timeline of digital finance evolution, 1971–2025. Events are color-coded by domain: electronic trading (blue), blockchain and DeFi (orange), AI/ML (green), regulation (purple), and major failures (red). The alternating above/below placement separates temporally proximate events. Sources compiled from Nar (????), Hendershott et al. (2011), Schär (2021), and regulatory databases.

1.1 Introduction

Between 2010 and 2025, the global financial system underwent a structural transformation more profound than any since the advent of electronic communications networks in the 1970s. Assets under management by robo-advisors surpassed USD 2 trillion. Daily cryptocurrency trading volumes routinely exceeded those of many national stock exchanges. High-frequency trading firms, operating on microsecond timescales, came to account for more than half of equity market volume in the United States and a growing share in Europe (Menkveld, 2013; Brogaard et al., 2014). Decentralized finance protocols, which did not exist before 2018, locked over USD 100 billion in smart contracts at

Joerg Osterrieder
Universiteit Twente, Enschede, The Netherlands

Bern Business School, Bern, Switzerland, e-mail: `j.r.osterrieder@utwente.nl`

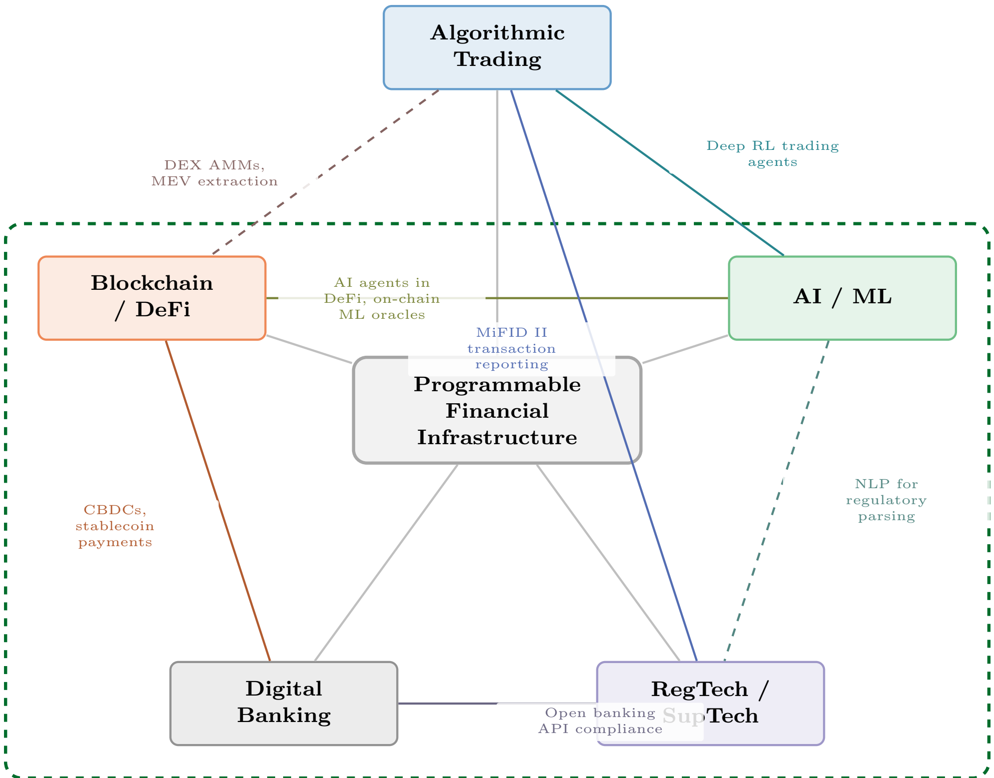

Fig. 1.1 Technology convergence map of digital finance. The five technology domains form an interconnected system whose pairwise interactions (labeled edges) generate capabilities absent in any single domain. The central thesis—programmable financial infrastructure—emerges from this convergence. The dashed green border highlights the convergence zone most relevant to sustainability applications, encompassing AI-driven ESG measurement, blockchain-based carbon markets, and digital banking's financial inclusion potential.

Table 1.1 Three-dimensional taxonomy of digital finance innovations. Cells contain representative examples coded by disruption level: **I** = incremental (improved components, unchanged architecture), **A** = architectural (reconfigured linkages), **D** = disruptive (new components and architecture). Empty cells indicate combinations with limited current activity. Classification follows Hen (????) adapted for financial services.

Financial Function	**AI / ML**	**Blockchain / DLT**	**Cloud / API**	**IoT**	**Big Data**
Payments & Settlement	**I** Fraud scoring	**D** CBDCs, stablecoins	**A** Open banking APIs (PSD2)	**I** Contactless NFC	**I** Real-time payment analytics
Lending & Credit	**A** Alt-data credit scoring	**D** DeFi lending (Aave, Compound)	**A** BaaS lending platforms	—	**A** Transaction-based underwriting
Investing & Trading	**D** Autonomous RL agents	**A** DEX AMMs, tokenized assets	**I** Cloud-based execution	—	**I** Satellite / alt-data alpha
Insurance & Risk Transfer	**I** Claims automation	**A** Parametric smart contracts	**I** API-based embedded insurance	**A** IoT telematics pricing	**I** Catastrophe modeling
Risk Management	**I** Stress testing, VaR	**I** On-chain risk monitoring	**I** Cloud risk engines	—	**A** NLP sentiment risk
Compliance	**A** NLP regulatory parsing	**A** On-chain KYC/AML	**A** RegTech API platforms	—	**I** Transaction surveillance

Note: Cell background indicates disruption level—light blue (I), light orange (A), light red (D)—following the Henderson–Clark typology.

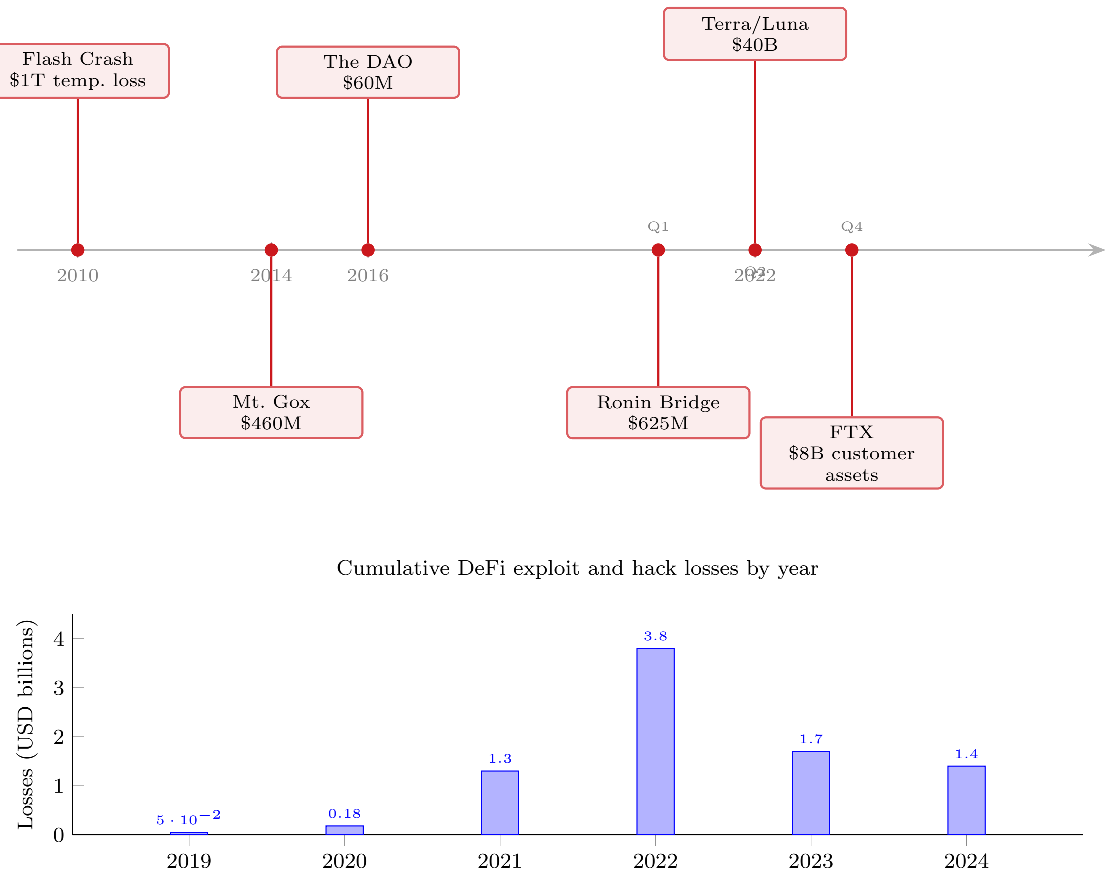

Fig. 1.2 Major digital finance failures and cumulative DeFi exploit losses. **Top:** Timeline of landmark failures with approximate financial impact. **Bottom:** Annual DeFi exploit and hack losses, showing the 2022 peak driven by the Ronin Bridge, Wormhole, and Nomad Bridge exploits alongside the Terra/Luna collapse. Data sources: Chainalysis (2024), DefiLlama, Commodity Futures Trading Commission enforcement records.

Table 1.2 ESG impact assessment matrix for five digital finance technology domains. Scores $T_{\text{ESG}} \in [-1, +1]$ represent directional assessments of net impact along Environmental (E), Social (S), and Governance (G) dimensions. Positive scores indicate net positive impact; negative scores indicate net negative impact. Scores are qualitative expert assessments informed by the literature cited in Section 1.6, not statistical estimates.

Technology Domain	**Environmental**		**Social**		**Governance**	
	T_{ESG}	Rationale	T_{ESG}	Rationale	T_{ESG}	Rationale
Algorithmic Trading	-0.2	Co-location energy	$+0.1$	Tighter spreads	-0.3	Flash crash risk
Blockchain / DeFi	-0.4	PoW energy; PoS mitigates	$+0.3$	Financial inclusion	-0.5	Governance gaps
AI / ML	-0.1	Training compute	-0.3	Bias, opacity	$+0.4$	Compliance automation
Digital Banking	$+0.2$	Paperless, branch reduction	$+0.6$	Inclusion, access	$+0.3$	Real-time oversight
RegTech / SupTech	$+0.1$	Efficient reporting	$+0.2$	Consumer protection	$+0.7$	Transparency, enforcement

Note: Scores reflect the assessment framework developed in Section 1.2.2. Environmental scores for blockchain assume current mix of PoW and PoS networks. Social scores for AI/ML reflect documented algorithmic bias in credit scoring (Bar, ????; Fuster et al., 2019). Governance scores for blockchain reflect the post-FTX regulatory landscape (FSB, ????a).

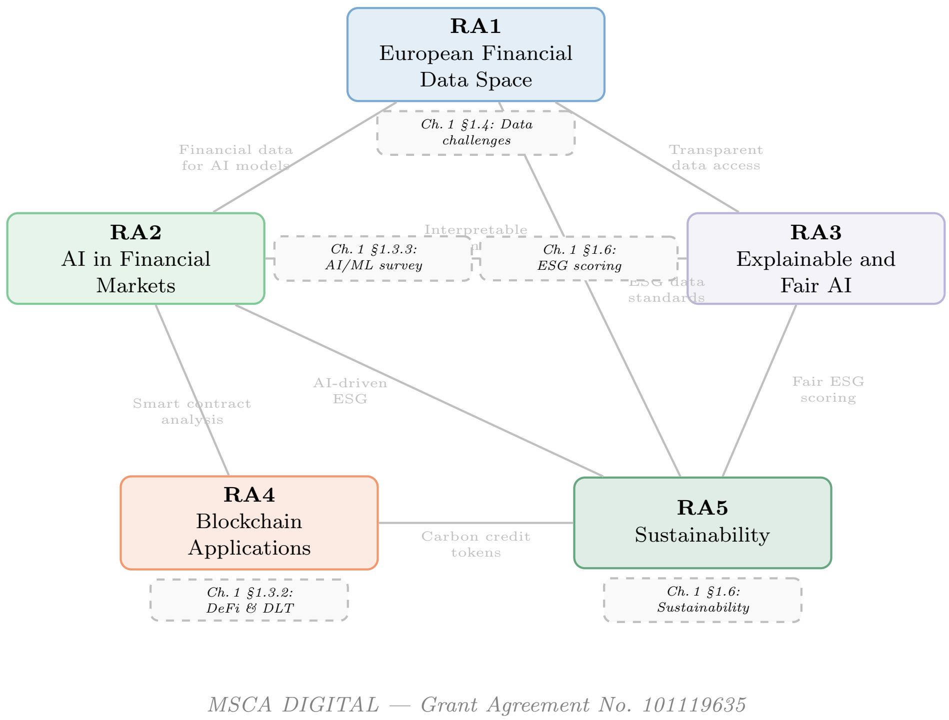

Fig. 1.3 The five research areas of the MSCA DIGITAL project and their interconnections. Dashed annotations map each research area to the corresponding section of Chapter 1. The network structure reflects the project's emphasis on cross-cutting themes: sustainability (RA5) connects to every other area, while AI (RA2) serves as the methodological backbone linking financial data infrastructure (RA1) to blockchain applications (RA4).

Table 1.3 Comparative overview of five digital finance technology domains across maturity, scale, regulatory status, sustainability impact, and primary research challenges as of 2025.

Domain	**Maturity**	**Market Size**	**Key Regulatory Framework**	**Sustainability**	**Primary Research Challenge**
Algorithmic Trading	Mature	>$12T daily	MiFID II, Reg NMS, Reg SCI	Mixed	Latency arms race welfare costs; AI agent market stability
Blockchain / DeFi	Growth	$2.5T market cap	MiCA, DORA, Travel Rule	Negative (improving)	Scalability trilemma; cross-chain security; MEV
AI / ML in Finance	Growth	$45B fintech AI	EU AI Act, SR 11-7	Mixed	Explainability–performance tradeoff; distributional shift
Digital Banking	Mature	3.6B users	PSD2/PSR, open banking mandates	Positive	Interoperability across jurisdictions; data portability
RegTech / SupTech	Early–Growth	$12B (est.)	Basel III reporting, AML6, DORA	Positive	Real-time supervision at scale; regulatory NLP accuracy

Note: Market size figures are approximate and source-dependent. Algorithmic trading volume reflects combined global equity, futures, and FX markets. Blockchain market cap is total cryptocurrency capitalization. AI market size is the financial-services AI segment. Digital banking user count is global mobile banking users. RegTech market size is estimated addressable market per industry reports.

Table 1.4 Summary of landmark digital finance failures. Events are ordered chronologically and analyzed by financial impact, root cause, and systemic lesson. Financial impact figures are approximate and reflect the most widely cited estimates in the academic and regulatory literature.

Event	**Date**	**Impact**	**Root Cause**	**Systemic Lesson**
Flash Crash	May 2010	\$1T[a]	Spoofing + algo feedback loops in thin liquidity	Algorithmic herding can amplify shocks beyond human intervention timescales; circuit breakers are necessary but insufficient
Mt. Gox	Feb 2014	\$460M	Insecure hot wallet management; no proof of reserves	Centralized crypto custody without segregation of client assets replicates pre-regulation banking risks
The DAO	Jun 2016	\$60M	Reentrancy vulnerability in Solidity smart contract	Smart contract code is law only until the code is wrong; immutability amplifies design flaws
Ronin Bridge	Mar 2022	\$625M	Compromise of 5/9 validator keys (state-sponsored attack)	Cross-chain bridges concentrate trust in small validator sets, creating single points of failure
Terra / Luna	May 2022	\$40B	Reflexive seigniorage mechanism without exogenous collateral	Algorithmic stablecoins lack a credible lender of last resort; endogenous collateral creates death spirals under stress
FTX	Nov 2022	\$8B[b]	Deliberate commingling of customer funds; no segregation	Vertical integration (exchange + custodian + market maker) without information barriers enables fraud at scale

[a]Temporary market value loss; recovered within minutes. [b]Customer asset shortfall as reported by the bankruptcy examiner (FTX, ????).

Sources: Kirilenko et al. (2017) (Flash Crash), Nar (????) (Mt. Gox), Mehar et al. (2019) (The DAO), Chainalysis (Ronin), Uhlig (2022) and Clements (2021) (Terra/Luna), FTX (????) (FTX).

their peak. Central banks in over 130 jurisdictions initiated digital currency research programs, with several moving to pilot deployments (Kosse and Mattei, 2023; Auer et al., 2024). This is not incremental change. It is the wholesale re-architecture of financial intermediation—who performs it, how it is governed, and on what technological substrate it operates. Gomber et al. (2017) and Lee and Shin (2017) provide comprehensive surveys of the digital finance and fintech ecosystems, respectively, mapping the business models, investment patterns, and technological architectures that drive this transformation.

The roots of this transformation extend back five decades. In the 1970s and 1980s, electronic trading began displacing floor-based exchange systems that had persisted for nearly two centuries. The New York Stock Exchange's adoption of the Designated Order Turnaround system in 1976, NASDAQ's shift to electronic market-making, and the London Stock Exchange's "Big Bang" of 1986 established the principle that markets are, fundamentally, information-processing systems (Kyle, 1985; Glo, ????). These early electronic markets retained human decision-making at their core—technology accelerated order routing and execution, but the strategies driving capital allocation remained manual. The efficiency gains were substantial: Philippon's analysis of the unit cost of financial intermediation shows that technology adoption was the primary driver of cost reduction in this era, even if aggregate intermediation costs remained stubbornly high relative to the scale of assets managed (Philippon, 2016; Bazot, 2018). Campbell et al. (2012) provide the foundational econometric methods for analyzing these market transformations, establishing the quantitative toolkit that subsequent empirical studies of electronic markets employ.

The 2000s brought a qualitative break. Algorithmic trading shifted decision-making from humans to machines operating at speeds and volumes that no human trader could match (Hendershott et al., 2011). High-frequency trading firms exploited latency advantages measured in microseconds, fundamentally altering market microstructure and price discovery (O'Hara, 2015; Budish et al., 2015). The empirical evidence on whether this transformation improved or degraded market quality remains contested: Hendershott et al. (2011) document liquidity improvements, while Budish et al. (2015) demonstrate that the resulting arms race constitutes pure rent-seeking. Aquilina et al. (2022) quantify the deadweight costs of latency competition at billions of dollars annually across global equity markets. What is unambiguous is that algorithmic trading transformed finance from a relationship-based industry into a technology-intensive one, setting the stage for every subsequent disruption.

The 2008 global financial crisis and the publication of the Bitcoin whitepaper (Nak, ????) initiated a third phase that challenged the institutional architecture of finance itself. Bitcoin proposed an alternative: a financial system without trusted intermediaries, secured instead by cryptographic proof and distributed consensus (Nar, ????). Ethereum extended this vision to programmable contracts, enabling financial logic to be encoded and executed without centralized counterparties (But, ????; Cong and He, 2019). The practical significance of these systems during the 2009–2015 period remained limited—transaction volumes were small, volatility was extreme (Osterrieder and Lorenz, 2017), and regulatory frameworks were nonexistent. But the conceptual contribution was profound: blockchain technology demonstrated that the separation of financial record-keeping from financial intermediation was technically feasible.

Between 2015 and 2020, artificial intelligence and machine learning moved from research curiosities to production systems in finance. Gu et al. (2020) demonstrated that machine learning models could generate economically significant improvements in asset pricing over linear factor models, establishing a new empirical benchmark. Deep learning architectures were deployed for credit scoring (Berg et al., 2020), derivatives hedging (Buehler et al., 2019), and textual analysis of corporate disclosures (Loughran and McDonald, 2011; Gentzkow et al., 2019). The arrival of natural language processing transformed how markets process information (Cao et al., 2023). This was not merely automation of existing processes but the creation of entirely new analytical capabilities: machine learning models identify non-linear patterns in high-dimensional financial data that traditional econometric approaches cannot capture (Israel et al., 2020; Dixon et al., 2020).

The period from 2020 to 2023 saw these previously separate technological trajectories begin to converge. Decentralized finance protocols replicated lending, trading, and insurance functions on blockchain infrastructure (Schär, 2021; Zetzsche et al., 2020), while automated market makers replaced traditional order books with algorithmic liquidity provision (Angeris et al., 2021). Tokenization extended blockchain-based ownership to traditional financial assets. Central bank digital currencies moved from conceptual proposals to active development programs (Auer et al., 2020; Brunnermeier et al., 2019; Agur et al., 2022). Embedded finance integrated financial services into non-financial platforms, blurring the boundary between technology firms and financial institutions (Stulz, 2019; Boot et al., 2021). The Global Findex Database (Demirgüç-Kunt et al., 2018) documents that 1.7 billion adults remained unbanked as of 2017, providing the demand-side motivation for much of the fintech innovation in payments and digital banking.

The most consequential developments, however, are those of 2024 and 2025. The U.S. Securities and Exchange Commission approved spot Bitcoin exchange-traded funds in January 2024, with BlackRock's iShares Bitcoin Trust attracting inflows that surpassed those of any ETF launch in history (Bla, ????; Dowling and Lucey, 2024). This approval represents institutional finance's definitive integration of cryptocurrency as an asset class, with implications for portfolio construction, custody infrastructure, and regulatory perimeter that are still unfolding (Li et al., 2024). The European Union brought the Markets in Crypto-Assets Regulation (MiCA) into full application (MiC, ????; ESM, ????), establishing the world's first comprehensive regulatory framework for crypto-assets. The Digital Operational Resilience Act (DORA) imposed technology risk management requirements across the financial sector (DOR, ????). The EU AI Act entered into force, introducing risk-based regulation of artificial intelligence systems used in credit decisions and financial surveillance (EUA, ????). Meanwhile, generative AI deployments in banking moved from experimentation to production: large language models are now used for regulatory reporting, client communication, and risk narrative generation (Wu et al., 2023; Lopez-Lira and Tang, 2024; Nie et al., 2024; Osterrieder, 2023). Real-world asset tokenization surged, with BlackRock's BUIDL fund and comparable instruments demonstrating that tokenized government securities can operate on public blockchain infrastructure (Nair et al., 2024).

Digital finance technologies are converging into a unified programmable financial infrastructure. This convergence—driven by AI, blockchain, and open data architectures—creates both unprecedented capacity for sustainable financial innovation and unprecedented systemic risks that current regulatory frameworks cannot adequately address. Data science serves as the critical bridge: it provides the analytical methods to measure and manage sustainability outcomes, to detect emerging systemic risks, to evaluate regulatory effectiveness, and to design financial systems that are simultaneously efficient, resilient, and aligned with societal objectives.

The case for a unified conceptual survey is compelling precisely because the boundaries between these technology domains are dissolving. AI models now execute on decentralized infrastructure. Tokenized assets require machine learning for valuation and risk management. RegTech systems depend on the same natural language processing capabilities that drive algorithmic trading strategies. Simultaneously, the European regulatory response—MiCA for crypto-assets, DORA for operational resilience, the AI Act for algorithmic decision-making, the EU Taxonomy and the Sustainable Finance Disclosure Regulation (SFDR) for sustainability disclosure (Arner et al., 2017; Buckley et al., 2020)—demands an integrated analytical framework. The sustainability imperative reinforces this need: the Corporate Sustainability Reporting Directive requires firms to report on environmental and social impacts using standardized metrics, while climate risk modeling demands the integration of physical and transition risk data into financial decision-making (Giglio et al., 2021; Krueger et al., 2020; Engle et al., 2020). Data science provides the tools to make these regulatory mandates operational—ESG measurement at scale, greenwashing detection through textual analysis, climate scenario modeling, and green bond verification through real-time monitoring (Berg et al., 2022; Bolton and Kacperczyk, 2021; Flammer, 2021). Without rigorous quantitative methods, sustainability in finance remains aspirational rather than actionable.

The remainder of this chapter is organized as follows. Section 1.2 develops a conceptual framework and taxonomy that organizes digital finance technologies along dimensions of automation, decentralization, and data intensity, drawing on innovation diffusion theory (Rog, ????) and architectural innovation models (Hen, ????). Section 1.3 surveys the five core technology domains: electronic and algorithmic trading, blockchain and distributed ledger technology, artificial intelligence and machine learning, decentralized finance and tokenization, and digital banking and payments. Section 1.4 examines data infrastructure and the European Financial Data Space, situating digital finance within the EU's data governance architecture. Section 1.5 confronts failures and systemic risks directly—the Terra/Luna collapse (Uhlig, 2022; Clements, 2021), the FTX fraud (FTX, ????), equity market flash crashes (Kirilenko et al., 2017)—analyzing how concentrated algorithmic strategies, opaque smart contracts, and fragmented oversight amplify tail risk. Section 1.6 treats sustainability as a standalone analytical dimension, examining measurement frameworks, technological enablers, and the tension between financial innovation and environmental impact. Section 1.7 maps open research questions onto the five thematic areas of the MSCA DIGITAL doctoral network, establishing the intellectual agenda that subsequent chapters in this volume develop. Section 1.8 concludes.

This chapter is a conceptual survey, not an empirical study. It synthesizes findings from the academic literature, regulatory documents, and industry developments through early 2026 to provide a structured foundation for the specialized treatments that follow. The chapter covers technologies and regulatory frameworks relevant to financial services in developed economies, with emphasis on European and North American markets. It does not address financial inclusion in developing economies, central bank monetary policy transmission, or payment system plumbing at the technical protocol level—each of these warrants dedicated treatment that lies beyond the scope of a single chapter. Where empirical evidence is cited, it is drawn from peer-reviewed publications or official reports by the BIS, ECB, FSB, and comparable institutions (BIS, ????a; FSB, ????a; IMF, ????).

1.2 A Conceptual Framework for Digital Finance

The digital finance landscape resists simple categorization. Technologies designed for one function migrate to others: machine learning models built for credit scoring are repurposed for fraud detection, then for regulatory reporting. Blockchain infrastructure conceived for payments now supports lending, insurance, and asset management. Any useful taxonomy must therefore capture not only what digital finance technologies *do*, but what technological substrate they operate on and how fundamentally they alter existing institutional arrangements. This section proposes a three-dimensional classification that organizes the field along functional, technological, and disruptive axes, then overlays sustainability considerations and maturity assessments onto that structure.

1.2.1 A Three-Dimensional Taxonomy

Hen (????) distinguished four types of innovation by decomposing change into its effects on components and on the linkages between them. Their framework—originally developed for photolithographic alignment equipment—maps onto financial services with surprising precision. An incremental innovation improves a component without altering the architecture: a faster execution algorithm within an existing exchange system, or a more accurate credit scoring model deployed within a traditional bank's lending pipeline (Fuster et al., 2019). An architectural innovation reconfigures the linkages between components while leaving individual components largely intact: open banking APIs that unbundle the relationship between deposit-taking and payment initiation (Boot et al., 2021), or DeFi protocols that replicate lending functions by composing smart contracts into novel configurations (Schär, 2021; Werner et al., 2022). A disruptive innovation—in the Christensen sense that Henderson and Clark's framework anticipates—replaces both components and architecture: central bank digital currencies that bypass the two-tier banking system entirely (Brunnermeier et al., 2019), or tokenized securities that eliminate custodial chains (Nair et al., 2024).

The first dimension of the proposed taxonomy captures *financial function*: the economic service that a technology performs. Six canonical functions span the landscape—payments and settlement, lending and credit, investing and trading, insurance and risk transfer, risk management and analytics, and regulatory compliance. These are not technology-specific categories; they reflect the economic purposes that financial systems serve regardless of their institutional form (Philippon, 2016; Thakor, 2020). A payment can be processed by a bank, a fintech, a blockchain protocol, or a mobile network operator. What matters for classification is the function, not the provider.

The second dimension captures *technology domain*: the computational and infrastructural substrate. Five domains dominate contemporary digital finance. Artificial intelligence and machine learning provide pattern recognition, prediction, and decision automation (Gu et al., 2020; Dixon et al., 2020). Blockchain and distributed ledger technology provide decentralized record-keeping, programmable contracts, and trustless settlement (Cong and He, 2019; Nar, ????). Cloud computing and open APIs provide scalable infrastructure and interoperability between systems (Stulz, 2019). The Internet of Things generates real-time data streams from physical assets—relevant for parametric insurance and supply chain finance. Big data analytics encompasses the statistical and computational methods applied to non-traditional data sources: satellite imagery for crop insurance, transaction graphs for anti-money laundering, and textual analysis of corporate disclosures (Gentzkow et al., 2019; Loughran and McDonald, 2011).

The third dimension captures *disruption level*, adapted from Hen (????): incremental enhancement (improved components, unchanged architecture), architectural innovation (reconfigured linkages, largely unchanged components), and disruptive transformation (new components, new architecture, new market structures). This dimension matters because the regulatory, competitive, and systemic implications differ qualitatively across levels. Incremental enhancements operate within existing regulatory perimeters. Architectural innovations strain those perimeters. Disruptive transformations render them obsolete.

Table 1.1 maps representative innovations across all three dimensions. The table is deliberately selective: the goal is not exhaustive enumeration but identification of the structural patterns that recur across function–technology–disruption combinations. Several observations emerge from the mapping. First, AI/ML penetrates every financial function, but at different disruption levels—incremental in risk management (better models within existing frameworks), architectural in lending (alternative data sources that restructure credit assessment), and potentially disruptive in trading (autonomous agents that bypass human decision-making entirely). Second, blockchain technology clusters at the architectural and disruptive levels; its comparative advantage lies not in doing existing things faster but in reconfiguring institutional relationships. Third, the combinations that pose the greatest regulatory challenge are those in the disruptive column, where existing supervisory frameworks lack jurisdiction, expertise, or both.

1.2.2 Sustainability Mapping

A taxonomy of digital finance that ignores sustainability is incomplete in 2026. The European Union's regulatory architecture—the Sustainable Finance Disclosure Regulation, the Corporate Sustainability Reporting Directive, the EU Taxonomy—demands that financial institutions measure and report sustainability impacts across their operations (Krueger et al., 2020; Giglio et al., 2021). The question is not whether digital finance technologies affect sustainability outcomes, but how to measure those effects rigorously enough to inform both investment decisions and regulatory policy.

Each technology–function combination in Table 1.1 generates sustainability externalities along three dimensions: environmental (energy consumption, carbon tracking, green product enablement), social (financial inclusion, algorithmic fairness, consumer protection), and governance (transparency, auditability, regulatory compliance). These externalities can be positive or negative, and they frequently conflict. Blockchain-based settlement eliminates intermediary risk and improves auditability—a governance benefit—while proof-of-work consensus mechanisms impose substantial energy costs—an environmental liability (Auer et al., 2022). AI-driven credit scoring expands access to underserved populations—a social benefit—while introducing algorithmic bias that may systematically disadvantage protected groups (Berg et al., 2020; Jagtiani and Lemieux, 2018). The net sustainability impact depends on implementation choices, not on the technology itself.

To formalize this assessment, define a scoring function $T_{\text{ESG}} : \mathcal{F} \times \mathcal{T} \times \mathcal{D} \times \{E, S, G\} \rightarrow [-1, +1]$ that maps each function–technology–disruption–ESG combination to a normalized impact score. This notation earns its place only if the methodology behind it is explicit. The scores in Table 1.2 are derived from a structured synthesis of the empirical literature rather than from subjective assignment. For environmental impacts, the primary sources are lifecycle energy analyses of blockchain networks (Auer et al., 2022), carbon emissions studies tied to financial portfolios (Bolton and Kacperczyk, 2021; Engle et al., 2020), and assessments of green bond effectiveness (Flammer, 2021). For social impacts, the evidence base includes studies of fintech lending to underserved populations (Jagtiani and Lemieux, 2018; Frost, 2020; Cornelli et al., 2023), algorithmic bias audits in credit markets (Berg et al., 2020), and analyses of financial inclusion through mobile money and digital payments (Arner et al., 2020). For governance, the sources are regulatory technology assessments (Arner et al., 2017; Buckley et al., 2020; Broeders and Prenio, 2018), transparency analyses of DeFi protocols (Aramonte et al., 2021; Werner et al., 2022), and studies of ESG disclosure quality (Berg et al., 2022).

Three limitations must be acknowledged. First, the T_{ESG} scores represent literature-weighted central tendencies, not precise measurements; different weighting schemes would produce different values. Berg et al. (2022) document that ESG ratings from major providers (MSCI, Sustainalytics, CDP, Refinitiv) correlate at only $\rho \approx 0.54$ on average—far below credit rating correlations of $\rho > 0.99$. Any technology-level ESG assessment inherits this measurement divergence. Second, the scores are time-dependent: Ethereum's transition from proof-of-work to proof-of-stake reduced its energy consumption by an estimated 99.95%, transforming its environmental score from strongly negative to approximately neutral. Third, the aggregation from technology-level scores to portfolio-level or system-level sustainability assessments is non-trivial. Pedersen et al. (2021) show that the relationship between ESG characteristics and expected returns is neither linear nor monotonic, and Avramov et al. (2022) demonstrate that ESG rating uncertainty itself constitutes a priced risk factor. These complications do not invalidate the scoring framework, but they constrain its appropriate use to directional assessment rather than cardinal measurement.

1.2.3 Technology Maturity and Adoption

The five technology domains occupy markedly different positions on the adoption trajectory, and the gap between institutional and retail uptake varies by domain in ways that carry direct implications for regulatory strategy and systemic risk.

Algorithmic trading and AI/ML in finance are the most mature domains. Algorithmic strategies account for over 60% of US equity volume and a growing share in European and Asian markets (Hendershott et al., 2011; Menkveld, 2013; Brogaard et al., 2014). Machine learning models are deployed in production for credit scoring, fraud detection, portfolio optimization, and regulatory reporting across the global banking sector (Gu et al., 2020; Israel et al., 2020). The release of large language models accelerated adoption further: by 2025, major banks had deployed generative AI for regulatory document analysis, client communication, and risk narrative generation (Wu et al., 2023; Lopez-Lira and Tang, 2024; Osterrieder, 2023). Institutional adoption is deep; retail exposure is largely indirect, mediated through robo-advisory platforms and AI-enhanced banking applications. The primary regulatory catalyst is the EU AI Act (EUA, ????), which imposes risk-based classification on AI systems used in credit decisions—a requirement that will force banks to document model governance, explainability, and bias testing at a level of granularity that most institutions have not previously maintained.

Blockchain and DLT occupy a more complex position. Institutional engagement accelerated sharply after the approval of spot Bitcoin ETFs in January 2024 (Bla, ????; Dowling and Lucey, 2024), and real-world asset tokenization demonstrated commercial viability with instruments like BlackRock's BUIDL fund (Nair et al., 2024). DeFi protocols, measured by total value locked, recovered from the post-2022 contraction but remained well below their November 2021 peak of approximately USD 180 billion (Werner et al., 2022; Aramonte et al., 2021). The BIS survey of central banks found that 93% were engaged in CBDC research, with 18 retail CBDC pilots active as of late 2023 (Kosse and Mattei, 2023; Auer et al., 2024). MiCA provides regulatory clarity in Europe (MiC, ????; ESM, ????), but the US regulatory environment remains fragmented across the SEC, CFTC, and state-level regulators (Goldstein et al., 2023). The gap between institutional and retail adoption is narrowing but structurally different: institutions engage primarily through custody, ETFs, and tokenization; retail users interact directly with wallets, exchanges, and DeFi protocols, bearing counterparty and smart contract risks that institutional participants largely avoid.

Cloud computing and open banking APIs are infrastructurally mature but regulatorily nascent. PSD2 mandated open banking in Europe; PSD3 will extend the framework to non-bank payment service providers. The adoption metrics here are less visible to end users—API call volumes, third-party provider registrations, embedded finance transaction values—but the structural impact is substantial (Boot et al., 2021; Buchak et al., 2018). Big tech platforms now originate significant lending volume in several markets, using data advantages that traditional banks cannot replicate (Cornelli et al., 2023; Frost, 2020).

IoT and parametric data-driven finance remain early-stage in financial applications, though insurance is the leading use case. Adoption is constrained by sensor deployment costs, data standardization challenges, and actuarial model recalibration requirements.

1.2.4 Geographic Variation

The taxonomy presented in Table 1.1 does not map uniformly across jurisdictions. Regulatory regimes, infrastructure endowments, and market structures produce sharply different digital finance landscapes, and these differences matter for any assessment of convergence patterns or sustainability outcomes.

The European Union has adopted the most comprehensive regulatory approach. MiCA provides a single passport regime for crypto-asset service providers (MiC, ????; ESM, ????). DORA imposes operational resilience requirements that cover ICT risk management, incident reporting, and third-party provider oversight (DOR, ????). The AI Act classifies AI systems used in credit decisions as high-risk, triggering mandatory conformity assessments (EUA, ????). PSD2 and PSD3 mandate open banking. The European Financial Data Space, examined in Section 1.4, extends this logic to financial data governance—creating a framework for cross-border data sharing that could enable pan-European AI model training and ESG analytics at a scale that fragmented national regimes cannot support.

The United States exhibits regulatory fragmentation. Crypto-asset supervision is contested between the SEC and CFTC, with classification disputes (security versus commodity) remaining unresolved for major tokens. Banking regulation is split across federal and state authorities. AI regulation is largely sector-specific rather than horizontal (Claessens et al., 2018). This fragmentation creates compliance complexity but also permits faster experimentation: the US remains the dominant jurisdiction for algorithmic trading innovation, venture-backed fintech, and—since January 2024—institutional crypto-asset products (Bla, ????).

Asia presents a bifurcated picture. China leads in CBDC deployment with the digital yuan (e-CNY) pilot spanning multiple provinces and hundreds of millions of users (Allen et al., 2022). South Korea and Japan have advanced regulatory frameworks for digital assets. Singapore functions as an innovation hub with relatively permissive sandbox regimes. India's Unified Payments Interface processed over 10 billion transactions monthly by 2024, representing the world's most successful real-time retail payment system—a cloud/API achievement with no blockchain component whatsoever (Frost, 2020).

Sub-Saharan Africa demonstrates that digital finance adoption need not follow the technology trajectory of developed markets. Mobile money—M-Pesa and its successors—achieved financial inclusion gains that traditional banking infrastructure could not deliver, processing transaction volumes that exceed the GDP of several African economies (Arner et al., 2020; BIS, ????a). The technology stack is relatively simple (USSD-based, not smartphone-dependent), the disruption level is unambiguously transformative, and the sustainability implications are overwhelmingly positive on the social dimension. This case underscores a point that the taxonomy makes visible: the most consequential digital finance innovations are not always the most technologically sophisticated.

These geographic variations carry a specific implication for the European Financial Data Space initiative discussed in Section 1.4. The EU's regulatory comprehensiveness creates a unique opportunity to build integrated data infrastructure that spans digital finance domains—but only if the data governance framework accommodates the cross-border, cross-technology data flows that the taxonomy identifies as structurally necessary. A framework designed around traditional banking data categories will fail to capture the DeFi transactions, on-chain analytics, and alternative data sources that the taxonomy shows are increasingly central to financial function delivery.

1.3 Technology Foundations and Literature

The taxonomy developed in Section 1.2 identifies five technology domains and maps them across financial functions, disruption levels, and sustainability dimensions. This section provides the evidence base for that mapping. Each subsection examines a technology domain through three lenses: the state of the academic literature, the frontier developments of 2024–2026, and the sustainability implications that the T_{ESG} scoring framework must capture. The organizing thesis is convergence: these domains are no longer separable research silos. Algorithmic trading now executes on decentralized infrastructure. AI models score the creditworthiness of DeFi borrowers. Blockchain settlement requires machine learning for compliance verification. Understanding each domain individually is necessary but insufficient; the interactions between them generate both the sustainability opportunities and the systemic risks that subsequent sections address.

1.3.1 Algorithmic Trading and Market Microstructure

The post-MiFID II European equity landscape operates under a transparency regime that fundamentally altered the economics of algorithmic trading. Double volume caps, systematic internaliser obligations, and consolidated tape requirements redistributed order flow from dark pools to lit venues and systematic internalisers, compelling proprietary trading firms to recalibrate strategies that had been optimized for the pre-2018 fragmentation pattern (Menkveld, 2013, 2016). In the United States, algorithmic strategies continue to account for over 60% of equity volume, with high-frequency market makers providing the majority of displayed liquidity (Hendershott et al., 2011; Brogaard et al., 2014). The empirical debate over whether this intermediation structure improves market quality has matured considerably since Hendershott et al. (2011)'s initial finding that algorithmic trading narrows spreads: Baron et al. (2019) document that HFT firms earn persistent and economically significant returns, while Aquilina et al. (2022) quantify the deadweight costs of the latency arms race at approximately USD 5 billion annually across global equity markets. The emerging consensus treats these findings as complementary rather than contradictory—algorithmic market makers do improve average liquidity, but the arms race to provide it faster extracts rents that exceed the liquidity benefits at the margin (Budish et al., 2015; Foucault et al., 2016). These dynamics reflect a broader structural property of financial markets: Chordia et al. (2000) demonstrate that liquidity exhibits strong commonality across securities, implying that algorithmic market-making strategies are exposed to systematic rather than idiosyncratic liquidity risk. Fama and French (2004) survey the theoretical and empirical status of the Capital Asset Pricing Model, whose equilibrium predictions algorithmic strategies both exploit and, through their aggregate behavior, help to enforce. Easley and O'Hara (2004) formalize the mechanism through which information asymmetry increases the cost of capital, providing the theoretical foundation for why algorithmic speed advantages translate into economic rents.

The analytical framework for optimal execution remains anchored in Almgren and Chriss (2001), whose contribution warrants formal treatment because its assumptions illuminate both the power and the limitations of analytical approaches. Consider an investor liquidating X shares over a time horizon $[0, T]$, partitioned into N intervals. Let x_k denote the remaining inventory at the start of interval k, and $n_k = x_{k-1} - x_k$ the number of shares traded in interval k. The Almgren-Chriss framework decomposes execution cost into two components: permanent price impact, which shifts the fundamental value proportionally to the cumulative quantity traded, and temporary price impact, which imposes a cost on individual trades proportional to trading rate. Under the critical assumption of *linear* temporary impact, the expected implementation shortfall takes the form

$$\mathbb{E}[\text{IS}] = \gamma \sum_{k=1}^{N} n_k + \eta \sum_{k=1}^{N} \frac{n_k^2}{\tau}, \tag{1.1}$$

where γ captures permanent impact (per-share cost that depends on trade direction), η captures temporary impact (per-share cost that depends on trading speed), and τ is the interval length. The variance of the execution cost is $\text{Var}[\text{IS}] = \sigma^2 \tau \sum_{k=1}^{N} x_k^2$, reflecting the risk from holding unexecuted inventory in a volatile market. Minimizing $\mathbb{E}[\text{IS}] + \lambda \text{Var}[\text{IS}]$ for a risk-aversion parameter λ yields a closed-form trading trajectory that interpolates between a uniform schedule ($\lambda = 0$) and immediate liquidation ($\lambda \to \infty$).

The insight that earns this framework its place in 2026 is not the closed-form solution itself but the *linearity assumption* on temporary impact. Bouchaud et al. (2009) established empirically that temporary price impact follows a concave (approximately square-root) function of trade size, not a linear one—a finding with profound implications for optimal scheduling. Cartea et al. (2015) extended the framework to algorithmic market making, deriving inventory-optimal quoting strategies under adverse selection risk that incorporate stochastic volatility and order flow toxicity. Guéant (2016) provided a comprehensive mathematical treatment bridging optimal execution and market making within a unified control-theoretic framework. The contemporary frontier moves further: reinforcement learning approaches to execution (Dixon et al., 2020) learn impact functions directly from data without parametric assumptions, but at the cost of interpretability and out-of-sample stability. The choice between analytical and ML-based execution thus reduces to a bias-variance tradeoff: Almgren-Chriss is wrong about the functional form of impact but generalizes reliably across market regimes; neural network execution policies achieve lower in-sample shortfall but suffer from overfitting to the specific microstructure regime on which they were trained (de Prado, 2018).

Machine learning signal generation in trading faces three fundamental challenges that the literature has not resolved. First, financial time series are non-stationary: the data-generating process changes as market participants adapt, invalidating the i.i.d. assumption that underlies standard supervised learning (Cont, 2011; Israel et al., 2020). Second, the signal-to-noise ratio is orders of magnitude lower than in the computer vision and NLP domains where deep learning achieved its canonical successes; Gu et al. (2020) report that even the best ML models explain only modest fractions of monthly return variation. Third, overfitting is structurally incentivized: the combinatorial space of potential trading features—Feng et al. (2020) document hundreds of proposed factors—combined with limited independent time-series observations creates a multiple testing problem that conventional cross-validation does not adequately address. Kozak et al. (2020) demonstrate that regularization (shrinkage toward the principal components of the return covariance matrix) substantially improves out-of-sample performance, suggesting that the factor zoo is largely a measurement artifact rather than evidence of market complexity.

The sustainability dimensions of algorithmic trading are quantitatively significant and intellectually underexplored. The energy footprint of HFT infrastructure—co-located servers, microwave networks, dedicated fiber-optic links—is concentrated among a small number of firms but imposes measurable environmental costs. Budish et al. (2015) estimate that latency competition generates private infrastructure spending of hundreds of millions of dollars annually, infrastructure whose sole purpose is to shave microseconds from execution times and whose energy consumption is wholly incremental to the financial system's operational needs. ESG signal-based trading strategies have proliferated since Bolton and Kacperczyk (2021) documented that carbon emissions predict stock returns: firms with higher Scope 1 and Scope 2 emissions earn higher average returns, consistent with a carbon risk premium. Whether algorithmic strategies can profitably exploit ESG signals remains contested. Pástor et al. (2021) show that sustainable investing in equilibrium generates a "green" asset pricing factor, but the returns to this factor depend critically on the rate of change in investor ESG preferences—a parameter that is difficult to estimate and unstable over time.

1.3.2 Blockchain, DLT, and Decentralized Finance

Blockchain technology's transition from a monetary experiment to programmable financial infrastructure is now well advanced. The analytical foundations were established by Cong and He (2019), who modeled how decentralized consensus resolves information asymmetries in financial contracting, and by Biais et al. (2019), who characterized the game-theoretic equilibria of proof-of-work mining. The formal representation of smart contract execution merits retention as notation. A smart contract operates as a deterministic state machine: given global state $\sigma_t \in \mathcal{S}$ and a transaction T containing instructions and input data, the state transition function Υ produces

$$\sigma_{t+1} = \Upsilon(\sigma_t, T), \tag{1.2}$$

where Υ is deterministic, publicly verifiable, and irreversible conditional on finality. This notation earns its presence because it makes precise what "programmable finance" means: financial contracts are not merely recorded on a shared ledger but *executed* by it, with the state transition function replacing the discretionary judgment of intermediaries (But, ????; Schär, 2021). The limitations are equally precise: Υ cannot access off-chain data without oracles, cannot enforce physical-world obligations, and executes with the gas costs and throughput constraints of the underlying consensus mechanism.

The post-Merge Ethereum landscape has reshaped the economic analysis of blockchain consensus. Saleh (2021) provided the theoretical foundations for proof-of-stake, demonstrating that validator incentives under staking produce consensus without the energy expenditure of proof-of-work. The practical consequence was dramatic: Ethereum's September 2022 transition reduced network energy consumption by an estimated 99.95%, transforming its T_{ESG} environmental score from strongly negative to approximately neutral (Auer et al., 2022). The post-Merge ecosystem, however, introduced new risks. Maximal Extractable Value (MEV)—the profit that block producers can capture by reordering, inserting, or censoring transactions (Daian et al., 2020)—persists under proof-of-stake and has spawned a secondary infrastructure of MEV relay networks, block builders, and searcher bots that concentrate value extraction in ways that undermine the decentralization thesis. The growth of liquid staking derivatives (Lido's stETH commands over 30% of staked ETH) and restaking protocols (EigenLayer) introduces leverage and rehypothecation risks to validator economics that the original proof-of-stake design did not anticipate (Werner et al., 2022). Li et al. (2017) provide a comprehensive survey of blockchain security threats, categorizing attacks on consensus mechanisms, smart contracts, and network layers that inform the risk analysis in Section 1.5. These are not implementation details; they represent a structural tension between the decentralization ethos and the economic logic of scale economies in validation.

Layer-2 scaling solutions—optimistic rollups (Arbitrum, Optimism) and zero-knowledge rollups (zkSync, StarkNet)—have partially resolved the throughput constraints that limited DeFi adoption, but at the cost of fragmenting liquidity across execution environments. The economics of L2 operation create natural oligopolies: sequencer revenue, which derives from ordering transactions within rollup batches, concentrates in entities that can amortize fixed infrastructure costs across large transaction volumes (Aramonte et al., 2021).

Decentralized finance protocols constitute the primary application layer for programmable financial contracts. The literature has matured from descriptive taxonomies to rigorous economic analysis. Automated market makers (AMMs) deserve particular attention because they represent a genuinely novel market design—one without precedent in traditional finance. The constant product formula ($x \cdot y = k$) that defined Uniswap v2 (Angeris et al., 2021) is now superseded; Uniswap v3 introduced concentrated liquidity, where liquidity providers allocate capital to specific price ranges $[p_a, p_b]$. Within a given range, the virtual reserves follow

$$L = \frac{\Delta y}{\Delta\sqrt{p}} = \frac{\Delta x}{\Delta(1/\sqrt{p})}, \tag{1.3}$$

where L is the liquidity parameter (constant within a tick range), p is the current price, and Δx, Δy denote the changes in token reserves. This piecewise construction achieves capital efficiency that approaches centralized limit order books for liquid pairs, but imposes impermanent loss on providers who must actively manage their price ranges—a requirement that favours sophisticated, often algorithmic, liquidity providers over passive participants (Lehar and Parlour, 2021; Park, 2021). Gudgeon et al. (2020) and Werner et al. (2022) provide comprehensive treatments of DeFi protocol economics, documenting the composability that allows protocols to be stacked (lending on top of AMM positions on top of staking derivatives) and the cascading liquidation risks that composability creates.

1.3.2.1 Stablecoins

Stablecoins occupy a structurally distinct position within the DeFi ecosystem because they serve as the unit of account and primary medium of exchange for on-chain activity. The distinction between fiat-backed stablecoins (USDT, USDC) and algorithmic stablecoins (the defunct UST) is not merely taxonomic but reflects fundamentally different risk architectures. Fiat-backed stablecoins are, in economic substance, narrow-bank deposits: their stability depends on reserve adequacy and redemption credibility. Griffin and Shams (2020) raised foundational questions about Tether's reserve transparency, and the concentration risk is stark—Tether and Circle together intermediate the majority of on-chain dollar-denominated activity (Aramonte et al., 2021). Algorithmic stablecoins attempted to maintain peg through endogenous mechanisms without external collateral. The Terra/Luna collapse of May 2022 (Uhlig, 2022) validated the theoretical critique of Clements (2021), who argued that reflexive seigniorage mechanisms are inherently fragile: the same feedback loop that maintains the peg during expansion (mint stablecoin, burn governance token) accelerates the collapse during contraction (burn stablecoin, mint governance token into a market with no buyers). The $40 billion destruction of value in 72 hours demonstrated that "algorithmic stability" was a contradiction in terms, at least under the designs deployed through 2022. Section 1.5 returns to this episode in analytical detail.

MiC (????) imposes reserve requirements, redemption rights, and governance standards on stablecoin issuers operating in the EU—the first comprehensive regulatory framework for these instruments. The FSB's recommendations for global stablecoin arrangements (FSB, ????b) address cross-border implications that MiCA's jurisdictional scope cannot fully capture.

The frontier developments of 2024–2026 in the blockchain domain centre on three phenomena. First, real-world asset (RWA) tokenization has moved from proof-of-concept to institutional deployment: BlackRock's BUIDL fund tokenizes US Treasury exposure on Ethereum, demonstrating that institutional-grade securities can settle on public blockchain infrastructure (Nair et al., 2024; Bla, ????). Second, institutional DeFi—permissioned instances of open-source protocols deployed for regulated counterparties—has emerged as a pragmatic bridge between decentralized technology and centralised compliance, though at the cost of the permissionless access that motivated DeFi's original design (Zetzsche et al., 2020; Capponi and Jia, 2021). Third, zero-knowledge proof technology has matured beyond scaling applications to enable privacy-preserving compliance: zkKYC protocols allow users to prove identity attributes (jurisdiction, accreditation status) without revealing underlying personal data, a construction that could resolve the GDPR–blockchain tension that has constrained European DeFi adoption (deF, ????).

On sustainability, the post-Merge energy reduction—99.95% by network-level estimates—eliminated blockchain's most damning environmental liability but did not resolve its sustainability challenges entirely. Carbon credit tokenization protocols (Toucan, KlimaDAO) demonstrated both the potential and the pitfalls of on-chain environmental markets: early implementations suffered from double-counting issues and the tokenization of low-quality, already-retired offset credits, leading to governance reforms that now require vintage verification and registry linkage (Flammer, 2021). The governance dimension of blockchain sustainability is more positive: on-chain transparency provides auditable records of financial flows that are structurally unavailable in traditional intermediation, enabling sustainability verification at a granularity that quarterly ESG reports cannot match (Yermack, 2017; Harvey et al., 2021).

1.3.3 AI and Machine Learning in Finance

The application of machine learning to financial problems has undergone a methodological transition. The first wave—documented comprehensively by Gu et al. (2020, 2021) and by Dixon et al. (2020)—established that neural networks, gradient-boosted trees, and regularized linear models could improve empirical asset pricing by learning non-linear interactions among firm characteristics. Chen et al. (2024) extended this programme to conditional factor models, demonstrating that deep learning architectures can capture time-varying factor loadings that traditional models assume constant. Jiang et al. (2023) applied convolutional networks directly to price chart images, achieving return predictability from visual patterns—a result that challenges efficient market theory at a conceptual level while remaining consistent with it at the risk-adjusted level (Fam, ????).

The architectural frontier has moved beyond feedforward and recurrent networks to models whose structural properties match the characteristics of financial data. Transformer architectures, which capture long-range dependencies through self-attention mechanisms, have shown promise for financial time series where relevant information may span irregular time intervals—earnings announcements, macro releases, geopolitical events—that fixed-window architectures cannot flexibly encode. State space models offer an alternative for sequence modelling that scales linearly with sequence length (versus the quadratic scaling of standard attention), making them computationally tractable for the ultra-high-frequency data that characterises market microstructure applications (Cont, 2011). Diffusion models, originally developed for image generation, have been adapted for synthetic financial data generation—producing realistic multi-asset return distributions that preserve cross-sectional correlations, tail dependencies, and volatility clustering in ways that parametric simulators cannot (Dixon et al., 2020). These synthetic datasets address the chronic data scarcity problem in finance: historical samples of market stress events are inherently small, and augmenting them with structurally faithful synthetic data can improve the robustness of risk models and stress testing frameworks.

The relationship between AI model performance and interpretability has moved from an academic debate to a regulatory mandate. The EU AI Act (EUA, ????) classifies AI systems used in credit scoring and insurance pricing as *high-risk*, triggering mandatory conformity assessments that include transparency requirements, human oversight provisions, and bias testing protocols. The ECB's guide on AI and machine learning models for internal ratings imposes documentation standards that require banks to demonstrate model interpretability—not merely as a best practice but as a supervisory expectation. The tension is real: Gu et al. (2020) demonstrate that non-linear models outperform their linear counterparts precisely because they capture interaction effects that linear specifications miss, yet explaining *why* a neural network assigns a particular credit score to a particular applicant remains technically difficult at the individual-prediction level. The regulatory response is not to ban complex models but to require that they be accompanied by interpretability layers—an approach that creates a market for explainable AI methods without constraining model architecture (Bartram et al., 2020). Whether post-hoc explanation methods faithfully represent model reasoning, rather than rationalising it, remains an open question (Sirignano et al., 2023).

1.3.3.1 Generative AI and Large Language Models

The deployment of large language models (LLMs) in finance constitutes the most rapid technology adoption cycle the industry has experienced. BloombergGPT (Wu et al., 2023) demonstrated that domain-specific pre-training on financial corpora improves performance on financial NLP tasks relative to general-purpose models, though subsequent open-source alternatives (FinGPT and its successors) challenged the assumption that proprietary data advantages are durable. Lopez-Lira and Tang (2024) documented that GPT-based sentiment scores predict stock returns, extending the text-as-data programme of Gentzkow et al. (2019) and the corporate disclosure analysis of Cao et al. (2023) to a model class that can process unstructured text without task-specific feature engineering. By 2025, major banks deployed LLMs for regulatory document parsing, client communication drafting, and risk narrative generation (Osterrieder, 2023; Nie et al., 2024). Autonomous AI agents operating in DeFi—executing arbitrage strategies, managing liquidity positions, and optimizing yield across protocols—represent the convergence of generative AI and decentralized finance, with implications for market structure that are not yet well understood theoretically.

AI's sustainability implications cut in multiple directions. On the positive side, NLP-based greenwashing detection has emerged as a measurable capability: trained classifiers can flag inconsistencies between corporate sustainability claims and actual emissions disclosures with precision that manual audit cannot match at scale (Berg et al., 2022; Gentzkow et al., 2019). AI-driven ESG scoring systems—deployed by MSCI, Sustainalytics, and their competitors—process satellite imagery, supply chain data, and textual disclosures to generate ESG ratings, but the evidence on whether these scores outperform simple heuristics (e.g., sector-based carbon intensity rankings) is mixed. Berg et al. (2022) document that ESG ratings from six major providers correlate at only $\rho \approx 0.54$, implying that the measurement methodology contributes as much variance as the underlying construct. On the negative side, the computational cost of training and deploying large language models is substantial: training a GPT-4-class model consumes energy equivalent to hundreds of households' annual consumption, and the inference costs of deploying these models at banking scale are non-trivial contributions to institutional carbon footprints (Giglio et al., 2021). On algorithmic fairness, the evidence is nuanced. AI credit scoring systems expand access to populations underserved by traditional banking—Jagtiani and Lemieux (2018) document that fintech lenders penetrate areas with limited bank branch coverage—but they also risk encoding historical discrimination patterns in their training data. The fairness metric literature identifies an impossibility result: demographic parity, equalized odds, and calibration cannot be simultaneously satisfied except in trivial cases (Berg et al., 2020). The EU AI Act's non-discrimination requirements for high-risk systems force institutions to make explicit choices among these metrics—a regulatory design that surfaces value judgments that were previously implicit in model development.

1.3.4 Digital Banking, Payments, and CBDCs

The neobank model, which attracted over USD 100 billion in venture capital funding between 2015 and 2022, has entered a period of reckoning. The fundamental unit economics challenge is straightforward: customer acquisition costs exceed lifetime revenue for the majority of digital-only banks that compete primarily on price and user experience rather than on proprietary financial products. Consolidation is accelerating—through acquisition, through regulatory-driven exits, and through the quieter process of neobanks adding traditional revenue lines (lending, insurance distribution) that erode their differentiation from incumbent banks (Buchak et al., 2018; Vives, 2019). The survivors are those that achieved sufficient scale to amortize fixed technology costs across large deposit bases, or those that identified defensible niches (SME banking, cross-border payments, embedded finance) where traditional banks' legacy infrastructure imposes structural cost disadvantages (Boot et al., 2021; Stulz, 2019).

Big tech entry into financial services represents a more durable competitive threat to traditional banking. Cornelli et al. (2023) document that big tech credit has grown to exceed fintech credit in several jurisdictions, driven by data advantages—transaction histories, social graphs, platform activity—that banks cannot replicate without equivalent platform ecosystems. The regulatory response has been asymmetric: the EU's PSD2 mandated that banks share customer data with authorised third parties, but imposed no reciprocal obligation on platform companies to share their data with banks or regulators (Frost, 2020). PSD3, expected to enter into force by 2026, partially addresses this asymmetry through expanded scope and interoperability requirements, but the fundamental data advantage of platform companies persists.

Central bank digital currencies represent the most consequential potential disruption to the two-tier banking system since the creation of deposit insurance. The theoretical literature has crystallized around several design questions: wholesale versus retail deployment, account-based versus token-based architecture, interest-bearing versus non-interest-bearing design, and the degree of intermediation (Brunnermeier et al., 2019; Keister and Sanches, 2023; Agur et al., 2022). Keister and Sanches (2023) demonstrate formally that a CBDC can improve payment efficiency while posing risks to bank funding through deposit substitution, with the net welfare effect depending on the CBDC's design parameters. Fernández-Villaverde et al. (2021) analyze the macroeconomic implications of "central banking for all," finding that a CBDC can improve financial inclusion and monetary policy transmission but may amplify bank runs if the CBDC serves as a frictionless safe haven during stress.

The ECB's digital euro project (ECB, ????) has advanced through its preparation phase with design decisions that reflect these trade-offs: offline functionality for privacy and resilience, holding limits to mitigate deposit substitution, and intermediation through supervised institutions rather than direct central bank accounts (Auer et al., 2024). The Federal Reserve's FedNow system, operational since July 2023, addresses the payments efficiency rationale for a CBDC through real-time gross settlement of existing bank deposits rather than through a new form of central bank money—a design choice that avoids the disintermediation risk but forgoes the programmability advantages of a native digital currency. The BIS mBridge project, linking central banks of China, Hong Kong, Thailand, and the UAE, tests the cross-border CBDC use case that domestic deployments cannot address (Kosse and Mattei, 2023; BIS, ????a).

The sustainability implications of digital banking and payments centre on financial inclusion. The World Bank's Global Findex database documents that approximately 1.4 billion adults remain unbanked globally, with the gap concentrated in Sub-Saharan Africa, South Asia, and among women in developing economies (Arner et al., 2020). Mobile money systems—M-Pesa and its successors—have demonstrated that digital payments infrastructure can achieve inclusion gains that traditional bank branch networks could not deliver, processing transaction volumes that exceed the GDP of several African economies (Frost, 2020; BIS, ????a). India's Unified Payments Interface processed over 10 billion transactions monthly by 2024 on a cloud/API stack with no blockchain component, representing the world's most successful real-time retail payment system. The technology substrate matters less than the institutional and regulatory design: M-Pesa operates on USSD protocols that predate smartphones; UPI operates on API standards built atop traditional bank rails. What both share is the regulatory willingness to permit non-bank entities to initiate payments, reducing the cost barrier that excluded the unbanked from formal financial services.

1.3.5 RegTech, SupTech, and Privacy

The volume and complexity of financial regulation have grown to a point where manual compliance is economically infeasible for all but the largest institutions. Arner et al. (2017) coined the term "RegTech" to describe the application of technology to regulatory compliance, distinguishing it from fintech by its orientation toward regulatory obligations rather than financial service delivery. The practical manifestation spans natural language processing for regulation parsing—extracting actionable requirements from legislative texts that span thousands of pages across multiple jurisdictions—automated transaction monitoring for anti-money laundering, and machine learning for fraud detection (Anagnostopoulos, 2018; Buckley et al., 2020). Supervisory technology (SupTech) mirrors these capabilities on the regulatory side: the BIS documents that central banks and financial supervisors in over 60 jurisdictions have deployed machine learning for supervisory purposes, including anomaly detection in prudential reporting data, network analysis of interbank exposures, and NLP for analysing regulated entities' public communications (Broeders and Prenio, 2018; BIS, ????b).

The intersection of data protection regulation and AI-driven financial services creates a tension that existing legal frameworks resolve imperfectly. The General Data Protection Regulation's (GDPR) right to explanation, data minimization principles, and purpose limitation constraints impose real constraints on the training data, model architectures, and deployment contexts available to financial AI systems (deF, ????). These constraints are not merely procedural: a credit scoring model trained on transaction data collected for payment processing purposes may violate purpose limitation even if the model itself is technically sound. The EU AI Act (EUA, ????) adds a second regulatory layer, requiring conformity assessments for high-risk AI systems that include bias testing, transparency documentation, and human oversight mechanisms—requirements that interact with GDPR obligations in ways that the regulatory texts do not fully harmonize.

Zero-knowledge proofs offer a cryptographic resolution to the privacy–compliance tension. A zero-knowledge proof allows a prover to convince a verifier that a statement is true without revealing any information beyond the truth of the statement itself. Applied to financial compliance, this enables constructions such as zkKYC, where a user proves to a DeFi protocol that they have been identity-verified by a regulated entity, that they are resident in a non-sanctioned jurisdiction, and that they meet accreditation thresholds—all without revealing their name, address, or identity documents (Schär, 2021). Polygon ID and similar implementations have moved these constructions from theoretical to deployable, though adoption remains limited by the computational overhead of proof generation and by the need for regulatory acceptance of cryptographic attestation as a substitute for traditional document verification.

The sustainability implications of RegTech are substantial and predominantly positive. Automated EU Taxonomy alignment tools can assess whether a financial product's underlying assets meet the technical screening criteria for environmentally sustainable activities—a classification task that the Taxonomy Regulation makes mandatory for large financial institutions but that manual analysis cannot perform at portfolio scale (Krueger et al., 2020; Giglio et al., 2021). ESG disclosure verification through NLP—comparing sustainability reports against emissions data, supply chain records, and satellite imagery—enables greenwashing detection that regulators currently lack the resources to perform systematically (Berg et al., 2022). The Corporate Sustainability Reporting Directive (CSRD), which extends mandatory sustainability reporting to approximately 50,000 EU companies, creates a compliance burden that will be met primarily through RegTech automation rather than through expansion of human compliance teams. The resulting datasets, if made interoperable through the European Financial Data Space, would constitute the most comprehensive sustainability information architecture ever constructed—a prospect examined in Section 1.4.

Cross-cutting synthesis. The five technology domains examined above are converging along three axes, and each convergence creates capabilities and risks that the individual domains do not exhibit in isolation. The AI–blockchain convergence is the most advanced: machine learning models now execute on-chain as autonomous agents, AI-driven analytics are applied to blockchain data for compliance and risk management, and generative AI accelerates smart contract development while simultaneously enabling more sophisticated exploit strategies (Werner et al., 2022; Nie et al., 2024). The AI–open banking convergence enables real-time credit decisions using transaction data shared through PSD2 APIs, personalized financial advice built on aggregated account data, and fraud detection systems that span institutional boundaries (Cornelli et al., 2023; Boot et al., 2021). The blockchain–payments convergence, manifested in stablecoins and CBDCs, challenges the correspondent banking model for cross-border settlement and creates alternative payment rails that operate outside the traditional banking perimeter (Auer et al., 2024; BIS, ????a).

Where convergence creates *new* risks, the mechanisms are specific and identifiable. AI agents operating on DeFi protocols can execute strategies at speeds that exceed the governance mechanisms designed to constrain protocol behavior, creating flash-loan-style exploits that combine ML sophistication with smart contract composability. Cross-protocol dependencies mean that a failure in one layer (stablecoin de-peg, oracle manipulation, validator censorship) propagates through interconnected systems in ways that single-domain risk models do not capture (Aramonte et al., 2023). The regulatory challenge is that each convergence crosses jurisdictional and institutional boundaries: AI regulation (the AI Act) does not account for on-chain execution contexts, crypto-asset regulation (MiCA) does not address AI-driven market manipulation, and data protection regulation (GDPR) was designed for centralized data controllers, not for decentralized protocols (FSB, ????a; IOS, ????).

Section 1.4 examines the data infrastructure that these converging technologies require and that the European Financial Data Space aims to provide. Section 1.5 returns to the risk dimension, analyzing specific failure episodes that expose the vulnerabilities created by cross-domain convergence.

1.4 Data Infrastructure and the European Financial Data Space

The convergence thesis articulated in Section 1.3 rests on an implicit premise: that the data required to train models, verify sustainability claims, and monitor systemic risk is available, interoperable, and of sufficient quality. This premise is far from satisfied. The financial data landscape is characterized by deep structural fragmentation, and the gap between the data that digital finance technologies *require* and the data that existing infrastructure *provides* constitutes the binding constraint on whether the convergence described in the preceding sections materializes as practice rather than aspiration.

1.4.1 The Financial Data Landscape

Traditional financial data infrastructure remains dominated by a small number of commercial vendors—Bloomberg, Refinitiv (now LSEG Data & Analytics), S&P Capital IQ, FactSet—whose pricing structures impose cost barriers that effectively exclude smaller institutions, academic researchers, and regulators from accessing the data necessary for independent analysis (Gol, ????). The consolidation of data provision mirrors the consolidation of the financial industry itself: LSEG's acquisition of Refinitiv for USD 27 billion and ICE's acquisition of Ellie Mae concentrated pricing power in an already oligopolistic market. Beyond cost, the standardization problem is severe. Vendor-specific identifiers (Bloomberg tickers, Refinitiv RICs, CUSIP, ISIN, SEDOL, FIGI) fragment the data space such that linking observations across providers requires non-trivial entity resolution—a problem that the Global Legal Entity Identifier Foundation (GLEIF) has partially addressed for counterparty identification through the LEI system, but that remains unsolved for instruments, transactions, and ESG attributes (GLE, ????).

The alternative data revolution has expanded the information set available to financial decision-makers beyond the traditional universe of prices, volumes, and accounting fundamentals. Satellite imagery of retail parking lots, container ship tracking data, credit card transaction aggregates, social media sentiment signals, web-scraped product reviews, and mobile application usage patterns now constitute a multi-billion-dollar industry (Gol, ????; Des, ????). The empirical evidence on whether alternative data generates economically significant alpha, net of acquisition costs, is mixed. Des (????) demonstrate that satellite-derived signals predict earnings surprises for the firms they cover, but the coverage itself is non-random—concentrated among large, visible companies in sectors amenable to physical observation—creating a selection bias that limits the generalizability of the findings. The deeper concern is that alternative data, by construction, favours institutions with the resources to acquire and process it, potentially widening the informational asymmetry between sophisticated and retail market participants (Zhu, ????).

The quality challenges confronting data-driven financial research are methodological as much as technological. Har (????a) demonstrated that the standard statistical thresholds applied in asset pricing research produce unacceptable false discovery rates when applied across the hundreds of factors that the literature has proposed, introducing multiple testing corrections that substantially reduce the number of factors that survive scrutiny. Survivorship bias—the systematic exclusion of failed firms, delisted securities, and discontinued funds from databases—inflates backtested returns in ways that practitioners frequently underestimate (de Prado, 2018). Look-ahead bias, where models are trained on information that would not have been available at the point of decision, remains endemic in machine learning applications to finance: Gu et al. (2020) impose strict temporal separation in their model evaluation, but a substantial portion of the ML-in-finance literature does not, rendering its reported performance metrics unreliable (Israel et al., 2020). Non-stationarity compounds these difficulties: the data-generating process in financial markets changes as participants adapt, such that a model trained on one regime may perform arbitrarily poorly in the next (Cont, 2011). For ESG data specifically, Berg et al. (2022)'s finding that major ESG rating providers correlate at only $\rho \approx 0.54$ implies that the measurement noise in sustainability data exceeds the signal in many applications—a conclusion with direct implications for any research programme that treats ESG scores as observed variables rather than noisy proxies.

Open data initiatives represent a structural counterweight to commercial data concentration. Pri (????), a fully open catalogue of the global research graph covering over 250 million scholarly works with metadata on authorship, citations, institutional affiliations, and topic classifications, enables bibliometric and scientometric analyses that were previously restricted to users of proprietary databases. In the financial domain, the EU Open Data Portal and the ECB's Statistical Data Warehouse provide free access to macroeconomic, monetary, and supervisory data, while GLEIF's open LEI database supports counterparty identification across jurisdictions (GLE, ????). These resources are necessary but not sufficient: open data solves the access problem but not the quality, timeliness, or granularity problems that define the frontier of financial data science.

1.4.2 The European Financial Data Space

The European Commission's strategy for a common European financial data space (EUD, ????a) represents the most ambitious attempt by any jurisdiction to construct interoperable data infrastructure for the financial sector through regulatory design rather than market evolution. The initiative connects directly to Research Area 1 of the MSCA DIGITAL network (Osterrieder et al., 2024), which investigates how data-driven methods can enhance financial risk assessment and sustainability measurement—investigations whose feasibility depends on the data access frameworks that the European Financial Data Space is designed to provide.

The legislative architecture rests on multiple interlocking instruments. The Financial Data Access Regulation (FiDA), proposed in June 2023, extends the open banking principles of PSD2 to insurance, pensions, investments, and creditworthiness data, requiring financial institutions to share customer-permissioned data through standardized APIs (EUF, ????). The Data Governance Act (EUD, ????b) establishes a framework for data intermediaries and introduces the concept of data altruism—voluntary data sharing for objectives of general interest, including scientific research and public policy—that could enable aggregated financial datasets for systemic risk monitoring without requiring institution-level disclosure. The Corporate Sustainability Reporting Directive (CSRD), which extends mandatory sustainability reporting to approximately 50,000 EU companies, generates the raw material for data-driven ESG analysis at a scale that voluntary reporting regimes never achieved (Krueger et al., 2020; Giglio et al., 2021).

The technical challenge of making these data frameworks operational while preserving privacy and confidentiality has catalysed a research programme at the intersection of cryptography and financial economics. Federated learning enables multiple institutions to train shared machine learning models without exposing their underlying data—a construction directly applicable to credit risk modelling, where banks hold complementary information about shared borrowers but face legal and competitive constraints on data pooling (Yan, ????). Lon (????) demonstrate that federated credit scoring models achieve performance comparable to centralized models trained on pooled data while satisfying the data minimization requirements of the GDPR, though the communication overhead and convergence guarantees remain active research areas. Secure multi-party computation allows regulators to compute aggregate statistics over institution-level data without observing individual submissions, enabling systemic risk monitoring that respects banking secrecy (Arc, ????). Homomorphic encryption, which permits computation on encrypted data without decryption, offers a more radical solution: regulatory reporting systems based on homomorphic schemes could allow supervisors to verify compliance properties of encrypted portfolios, though the computational costs of fully homomorphic encryption remain prohibitive for all but the simplest operations at current performance levels (Arm, ????).

The tension between these data-sharing ambitions and data protection obligations reflects a deeper divergence in regulatory philosophy across jurisdictions. The European model, anchored in GDPR and extended through FiDA and the Data Governance Act, prioritizes individual data rights and institutional consent mechanisms, creating a framework that is privacy-preserving by design but imposes friction on the data flows that innovation requires (deF, ????; Arner et al., 2017). The US approach relies on market-driven data aggregation through commercial vendors and voluntary industry standards, producing higher data availability for well-resourced participants but deeper access inequalities and limited regulatory visibility into data usage patterns (Gol, ????). China's model, exemplified by the social credit system and the centralized data architectures underlying Alipay and WeChat Pay, maximizes data availability for state-directed objectives but at costs to individual privacy that the European framework explicitly rejects (Allen et al., 2022; Frost, 2020). Whether the European model can achieve sufficient data fluidity to support competitive innovation while maintaining its privacy commitments is an empirical question that the next five years of FiDA and EFDS implementation will begin to answer.

1.4.3 Reproducibility and Open Science Infrastructure

The credibility of data science applied to finance depends on whether reported results can be independently verified. The evidence on this point is discouraging. Hou (????) attempted to replicate 452 anomalies documented in the asset pricing literature and found that the majority—roughly 65%—fail to meet conventional significance thresholds when tested on independent samples with consistent methodological standards. Har (????a) attribute much of this non-replicability to multiple testing: the cumulative body of factor research represents hundreds of correlated hypothesis tests, and conventional p-values, which do not account for the breadth of the search, overstate the evidence for individual factors. In machine learning applications, the replication challenge is compounded by sensitivity to hyperparameters, random seeds, data preprocessing choices, and the specific temporal windows used for training and evaluation (de Prado, 2018; Dixon et al., 2020).

The Quantlet and Quantinar platforms (Har, ????b,c) address these challenges by providing infrastructure for executable, versioned research code linked to standardized datasets and documentation. Quantlet serves as a repository of validated statistical and econometric code snippets—each associated with a specific publication, dataset, and computational environment—enabling line-by-line replication of published results. Quantinar extends this model to a collaborative teaching and research platform where quantitative methods, including those applied throughout this volume, can be demonstrated with live code execution against curated financial datasets. The companion materials for this book, hosted on these platforms, allow readers to reproduce the analyses presented in each chapter and to extend them with alternative specifications, updated data, or modified parameters.

The reproducibility imperative is particularly acute for sustainability claims. When a financial product is marketed as Paris-aligned or taxonomy-compliant, the underlying analysis—emissions calculations, scenario models, alignment metrics—must be independently verifiable to distinguish genuine green finance from greenwashing (Berg et al., 2022; Flammer, 2021). The CSRD's requirement for limited assurance on sustainability reports creates demand for reproducible analytical pipelines that auditors can evaluate, and the extension to reasonable assurance (anticipated by 2028) will intensify this demand further. Open science infrastructure is thus a governance mechanism: by making analytical methods transparent and results replicable, it constrains the interpretive latitude that enables greenwashing and supports the credibility of the sustainability data architecture that the European Financial Data Space aims to construct.

Section 1.5 turns from the infrastructure that enables digital finance to the episodes in which that infrastructure has failed, examining specific crises that expose the vulnerabilities created by cross-domain convergence and insufficient data governance.

1.5 Risks, Failures, and Systemic Concerns

The preceding sections have traced the convergence of algorithmic trading, blockchain infrastructure, artificial intelligence, digital banking, and regulatory technology into an increasingly interconnected digital finance ecosystem. That convergence generates capabilities that no single domain could produce in isolation, but it also creates fragility of a kind that traditional financial risk models were not designed to capture. The same composability that enables a DeFi lending protocol to accept staked derivatives as collateral, price them through an automated market maker, and hedge the resulting exposure through an on-chain options protocol also means that a failure at any point in the chain propagates through every layer simultaneously. This section examines the empirical evidence from major failures, identifies the systemic risk channels that digital finance has opened, and confronts the governance and accountability gaps that existing regulatory frameworks have not yet closed. The precedent for understanding how interconnected financial systems amplify shocks is well established: Brunnermeier (2009) provides the canonical analysis of the 2007–2008 liquidity and credit crunch, demonstrating the loss spiral and margin spiral mechanisms that digital finance ecosystems can replicate at algorithmic speed.

1.5.1 Landmark Failures and Their Root Causes

The collapse of the Terra/Luna ecosystem in May 2022 destroyed approximately $40 billion in value within 72 hours and exposed a design flaw that theoretical work had identified before the failure occurred. Clements (2021) argued that algorithmic stablecoins relying on reflexive seigniorage mechanisms are inherently fragile: the mint-and-burn arbitrage loop that maintains the peg during expansion—users burn LUNA to mint UST when demand rises, absorbing supply pressure through the governance token—operates in precise reverse during contraction, flooding the market with LUNA into an environment where no buyer of last resort exists. Uhlig (2022) formalized the death spiral dynamics, showing that once UST redemptions exceeded the market's capacity to absorb newly minted LUNA, the governance token price entered a reflexive collapse that fed back into further UST depegging, eliminating any rational basis for maintaining peg confidence. The root cause was not a technical exploit but an economic design that substituted endogenous token demand for exogenous collateral—a substitution that worked only so long as aggregate demand for the ecosystem's yield products, principally Anchor Protocol's 19.5% deposit rate, exceeded redemption pressure.

The contagion chain that followed was more instructive than the initial collapse. Celsius Network, which had deployed substantial client assets into Anchor's yield strategy, faced a liquidity crisis when Terra's collapse eliminated its largest yield source while simultaneously triggering client withdrawals. Three Arrows Capital, a crypto hedge fund with leveraged long positions across DeFi protocols, became insolvent as correlated asset price declines produced margin calls it could not meet. The bankruptcies of Voyager Digital and BlockFi followed through direct counterparty exposure to Three Arrows. Each link in this chain involved a different institutional form—algorithmic stablecoin, centralized yield platform, hedge fund, retail brokerage—but the common mechanism was maturity and liquidity transformation without adequate reserves, precisely the fragility that banking regulation was designed to prevent in the traditional financial system (Aramonte et al., 2023; FSB, ????a). The absence of circuit breakers, mandatory disclosure of counterparty exposures, or lender-of-last-resort facilities meant that contagion proceeded until the capital was exhausted. The episode revealed that DeFi's claim to "trustlessness" was accurate only at the protocol execution layer; the ecosystem's actual operation depended on a dense network of trust relationships among centralized intermediaries that operated without the regulatory safeguards governing their traditional finance counterparts.

The FTX fraud, which culminated in the exchange's bankruptcy in November 2022, was categorically different from Terra's collapse but exposed equally fundamental structural problems. The bankruptcy examiner's report (FTX, ????) documented that FTX commingled customer funds with its affiliated trading firm Alameda Research, used customer deposits to fund venture investments and political donations, and maintained accounting systems so deficient that the exact magnitude of misappropriation remained uncertain months after the filing. The fraud itself was straightforward—misappropriation of customer assets—and existing securities law, had it been applied, would have prohibited every element: commingling of customer and proprietary funds, self-dealing between affiliated entities, and operation of an exchange, custodian, market maker, and proprietary trading firm as a single entity without information barriers. The structural lesson is that FTX's regulatory arbitrage—incorporating in the

Bahamas under a regime that lacked the supervisory capacity to enforce its own licensing requirements—was a feature of the crypto industry's regulatory landscape, not an aberration. FTX was, at the time of its collapse, the third-largest cryptocurrency exchange globally, had secured investments from Sequoia Capital, BlackRock, and the Ontario Teachers' Pension Plan, and maintained regulatory licenses or registrations in multiple jurisdictions (IOS, ????; FSB, ????a). The failure of institutional due diligence at this scale suggests that the problem extended beyond one fraudulent actor to an industry structure where exchanges routinely combined functions that securities regulation has kept separate since the 1930s.

Flash crashes illustrate a different failure mode: the emergent behavior of algorithmic systems under stress conditions that their designers did not anticipate. The May 6, 2010 flash crash, during which the Dow Jones Industrial Average fell approximately 9% in minutes before recovering, was traced by Kirilenko et al. (2017) to the interaction between a large sell order executed by an algorithm that was insensitive to price and a population of high-frequency trading firms that shifted from liquidity provision to liquidity consumption as volatility exceeded their risk limits. The August 2015 ETF flash crash, during which exchange-traded funds traded at discounts of 30–40% to their net asset values, revealed that the arbitrage mechanisms connecting ETF prices to underlying basket values break down when market makers withdraw from both markets simultaneously. Currency flash crashes in January 2019, which saw the Japanese yen appreciate 4% against the US dollar in minutes during low-liquidity Asian trading hours, demonstrated that the phenomenon is not confined to equity markets (Aquilina et al., 2022; Budish et al., 2015). Circuit breakers—trading halts triggered by price movements exceeding predetermined thresholds—have been refined after each episode, but they address symptoms rather than causes. The fundamental problem is that algorithmic market makers provide liquidity as a profit-maximizing activity, not as a public utility, and their rational response to extreme volatility is to withdraw precisely when liquidity is most needed (Menkveld, 2013; Baron et al., 2019).

DeFi-specific exploits constitute a fourth category of failure, distinguished by their exploitation of smart contract logic or cross-protocol interactions rather than market dynamics. Bridge hacks—the Ronin Network breach ($625 million, March 2022) and the Wormhole exploit ($320 million, February 2022)—targeted the cryptographic infrastructure connecting separate blockchain networks, where validator key compromise or signature verification vulnerabilities enabled attackers to mint unbacked tokens on the destination chain. Smart contract vulnerabilities, including reentrancy attacks, oracle manipulation, and flash loan exploits, have generated cumulative losses exceeding $6 billion through 2024 across DeFi protocols (Werner et al., 2022; Gudgeon et al., 2020). Flash loan attacks deserve particular attention because they represent a genuinely novel attack vector: by borrowing and repaying within a single atomic transaction, an attacker can manipulate prices across multiple protocols, extract value through the manipulated price, and repay the loan—all without committing any capital. The persistence of these exploits despite formal verification tools and professional audit firms reflects a fundamental challenge: the composability that defines DeFi means that each protocol's security depends not only on its own code but on the behavior of every protocol it interacts with, a dependency surface that grows combinatorially with ecosystem complexity (Daian et al., 2020; Schär, 2021).

1.5.2 Systemic Risk Channels in Digital Finance

The composability that Section 1.3.2 identified as DeFi's defining architectural feature creates systemic risk through specific, identifiable contagion pathways. The most consequential pathway runs through stablecoins: Tether (USDT) and Circle (USDC) together intermediate the majority of on-chain dollar-denominated activity, and a loss of confidence in either issuer's reserves would trigger simultaneous liquidations across every DeFi protocol that uses these tokens as collateral, unit of account, or settlement medium (Griffin and Shams, 2020; Aramonte et al., 2021). Unlike a traditional bank run, where deposit insurance and central bank lending facilities provide backstops, a stablecoin run would propagate at the speed of blockchain finality—seconds to minutes—through automated liquidation mechanisms that cannot distinguish between fundamental insolvency and temporary illiquidity.

The crypto–traditional finance linkage has strengthened substantially since 2024. The approval of spot Bitcoin ETFs in the United States (Bla, ????; Dowling and Lucey, 2024) created a direct transmission channel between cryptocurrency price volatility and traditional investment portfolios, including retirement accounts. Bank exposure to crypto-assets, while still modest in aggregate, is concentrated among institutions that serve as custodians, prime brokers, or settlement providers for the crypto industry (Auer et al., 2022; Allen et al., 2022). Correlated liquidation cascades represent the most operationally dangerous contagion mechanism: when DeFi lending protocols automatically liquidate undercollateralized positions, the liquidation sales depress collateral prices, which triggers further liquidations in a feedback loop that the Terra/Luna episode demonstrated at scale (Gudgeon et al., 2020; Werner et al., 2022).

Concentration risk in the crypto ecosystem contradicts the decentralization thesis articulated by its proponents. Binance processes a plurality of global cryptocurrency trading volume. Tether's dominance of the stablecoin market creates a single point of failure for on-chain dollar liquidity. Ethereum validator concentration—with Lido's liquid staking protocol controlling over 30% of staked ETH and the top four staking providers collectively exceeding 60%—means that proof-of-stake's security guarantees depend on a small number of entities whose coordinated failure or regulatory seizure could compromise the network's consensus mechanism (Aramonte et al., 2021, 2023). These are too-big-to-fail dynamics emerging in a system designed to eliminate too-big-to-fail institutions.

Cyber risk adds a dimension that is specific to digital infrastructure. The European Systemic Risk Board (ESR, ????) identified systemic cyber risk—the possibility that a successful attack on a critical financial infrastructure node could cascade through interconnected systems—as a macroprudential concern distinct from individual institution-level cybersecurity. The Digital Operational Resilience Act (DORA) (DOR, ????), which entered into application in January 2025, imposes ICT risk management, incident reporting, and third-party provider oversight requirements on EU financial entities, but the regulatory perimeter does not extend to decentralized protocols or to the cloud infrastructure providers on which an increasing share of financial services depend. Ransomware attacks targeting financial institutions have increased in frequency and sophistication, with the operational disruption costs often exceeding the ransom demands themselves (BIS, ????b; Buckley et al., 2020).

1.5.3 Governance, Accountability, and Emerging Threats

The governance architecture of decentralized finance creates accountability gaps that existing legal frameworks struggle to address. When a DAO's smart contract fails—through a bug, an exploit, or an economic design flaw—the question of legal liability has no straightforward answer. The DAO's token holders may have voted on the contract's deployment, but decentralized governance through pseudonymous token voting does not create the fiduciary duties, corporate officer liability, or regulatory accountability that attach to the directors and officers of regulated financial institutions. deF (????) documented this tension between the "rule of code" and the rule of law; the subsequent years have not resolved it. MiC (????) addresses the problem for identifiable crypto-asset service providers by imposing governance and liability requirements, but fully decentralized protocols—those without identifiable operators—remain outside MiCA's regulatory perimeter, a gap that IOS (????) has identified as a priority for international coordination.

Quantum computing poses a medium-term threat to the cryptographic foundations of both traditional and decentralized finance. The security of RSA and elliptic curve cryptography (ECDSA), which underpin TLS connections to banking infrastructure and the signature schemes securing blockchain transactions, rests on the computational intractability of integer factorization and discrete logarithm problems—intractability that Shor's algorithm eliminates given a sufficiently powerful quantum computer. NIST's post-quantum cryptography standardization process, finalized in 2024 with the selection of CRYSTALS-Kyber for key encapsulation and CRYSTALS-Dilithium for digital signatures, provides the replacement algorithms, but the migration challenge is formidable: every TLS certificate, every hardware security module, and every blockchain address must transition to quantum-resistant schemes before cryptographically relevant quantum computers become operational (Nar, ????). Estimates of this timeline vary widely—from the early 2030s to beyond 2040—but the "harvest now, decrypt later" threat, in which adversaries collect encrypted data today for future decryption, means that migration must begin well before the threat materializes. For blockchain systems, the challenge is compounded by the immutability property: funds secured by ECDSA signatures on addresses whose public keys have been revealed (through prior transactions) cannot be retroactively protected without a coordinated network upgrade.

AI model risk in finance extends beyond the interpretability concerns discussed in Section 1.3.3 to encompass systemic vulnerabilities. Model monoculture—the deployment of similar or identical machine learning models across multiple institutions for credit scoring, fraud detection, or trading signal generation—creates correlated failure modes: if the same model architecture produces similar errors under the same market conditions, the resulting actions (simultaneous position liquidation, coordinated credit tightening, synchronized fraud alert failures) amplify rather than diversify risk (Bartram et al., 2020; Sirignano et al., 2023). Adversarial attacks on trading algorithms, demonstrated in controlled settings, show that small, carefully crafted perturbations to market data inputs can cause models to take positions that benefit the attacker (Dixon et al., 2020; Gu et al., 2020). The EU AI Act's (EUA, ????) classification of financial AI systems as high-risk and its requirement for conformity assessments, human oversight, and bias testing represent the first regulatory response to these concerns, but the Act's provisions were designed for centralized AI deployments, not for autonomous AI agents operating on decentralized protocols where no identifiable deployer bears regulatory responsibility.

The failures and risks catalogued in this section are not arguments against digital finance but arguments for its maturation. Every failure examined here has a structural explanation and, in most cases, a regulatory or design remedy that is either already enacted (MiCA, DORA, the AI Act) or under active development (post-quantum migration, DeFi governance frameworks). The pattern across all four categories—algorithmic instability, fraud enabled by regulatory arbitrage, flash crashes from liquidity withdrawal, and smart contract exploits from compositional complexity—is that digital finance systems fail when they replicate the functions of traditional financial intermediation without replicating its safeguards. The sustainability argument advanced throughout this chapter depends on this recognition: the convergence of technologies examined in the preceding sections creates genuine opportunities for financial inclusion, transparency, and efficiency, but only if the systemic risks that convergence simultaneously creates are identified, measured, and governed with the same rigor that the opportunities are pursued.

1.6 The Sustainability Dimension

A book titled *Bridging Data Science and Sustainable Financial Innovation* must confront a prior question: what, precisely, makes sustainable finance so difficult that it requires data science at all? The answer lies in measurement. Traditional financial quantities—prices, returns, cash flows, default probabilities—are observed with high frequency, reported under standardized accounting rules, and verified through audit. ESG characteristics possess none of these properties. They are measured infrequently, reported under competing and incompatible frameworks, and verified—when they are verified at all—through processes whose reliability is unknown. This measurement gap is not a technical inconvenience. It is the central obstacle to integrating sustainability into financial decision-making, and its resolution is the core argument of this volume: data science does not merely improve sustainability measurement; it makes evidence-based sustainable finance possible for the first time.

1.6.1 The Measurement Problem

The scale of ESG data divergence is empirically documented and striking. Berg et al. (2022) analyze ESG ratings from six major providers (MSCI, Sustainalytics, Moody's, S&P Global, CDP, and Bloomberg) and find average pairwise correlations of $\rho \approx 0.54$—a figure that would be alarming for any financial metric and that stands in stark contrast to credit rating correlations exceeding $\rho = 0.99$. The divergence is not random noise. Their decomposition identifies three sources: scope divergence (providers measure different attributes), measurement divergence (providers measure the same attribute differently), and weight divergence (providers assign different importance to the same attribute). Scope divergence accounts for the largest share, meaning that the ESG ratings industry has not converged on what sustainability *is*, let alone how to quantify it. Cha (????) document that this disagreement is not merely academic—firms that receive high ESG ratings from one provider frequently receive low ratings from another, creating a selection problem for investors who rely on these scores. The T_{ESG} scoring function introduced in Section 1.2.2 explicitly inherits this measurement uncertainty: the directional assessments it provides are constrained by the same input data that produces divergent ratings at the firm level. Drempetic et al. (2019) document a systematic firm-size bias in ESG scores—larger firms receive higher ratings partly because they have more resources for disclosure rather than better sustainability performance—adding a further layer of measurement noise that data science methods must address.

This divergence has direct financial consequences. Avramov et al. (2022) demonstrate that ESG rating disagreement constitutes a priced risk factor: stocks with high rating dispersion earn lower risk-adjusted returns, suggesting that investors treat measurement uncertainty itself as a source of risk. Gibson Gib (????) confirm this finding using institutional investor portfolio data, showing that disagreement reduces the effectiveness of ESG integration strategies. If the inputs to sustainable finance are unreliable, the outputs—portfolio tilts, engagement decisions, regulatory compliance assessments—inherit that unreliability.

Data science offers partial resolution through three channels. First, natural language processing enables systematic analysis of corporate sustainability reports, moving beyond the self-reported metrics that rating agencies rely upon. Bin (????) develop a climate-specific NLP framework that distinguishes between forward-looking commitments and backward-looking disclosures in corporate reports, identifying a "commitment–action gap" in which firms' stated targets substantially exceed their implemented measures. Luc (????) extend this approach using transformer-based models to classify climate claims against the EU Taxonomy's technical screening criteria, providing automated verification at a scale that manual audit cannot achieve. These methods do not eliminate greenwashing, but they make it detectable.

Second, satellite and geospatial data provide independent verification of environmental claims. Satellite-based measurement of methane emissions, deforestation rates, and industrial activity creates an observational layer that is orthogonal to corporate self-reporting (Giglio et al., 2021; Bolton and Kacperczyk, 2021). When a company reports declining emissions but satellite data shows stable or increasing thermal signatures from its facilities, the discrepancy is informative regardless of which measurement is correct. The combination of NLP-based report analysis and satellite-based physical verification creates a form of cross-validation that no single data source can provide.

Third, machine learning approaches to ESG score construction attempt to resolve divergence by learning composite scores from multiple providers. The question is whether this aggregation reduces noise or merely averages biases. Berg et al. (2022) are cautious on this point, arguing that aggregation without addressing scope divergence produces scores that are precise but not necessarily accurate. The evidence to date suggests that ML-based ESG scoring works best when the task is narrowly defined—predicting carbon emissions from financial statements, for instance, rather than constructing holistic sustainability assessments (Bolton and Kacperczyk, 2021; Engle et al., 2020).

Climate risk modeling represents the most technically mature application of data science to sustainability measurement. Physical risk assessment combines asset-level location data with climate projection models to estimate the probability and severity of damage from extreme weather events, sea-level rise, and temperature-related productivity losses (Giglio et al., 2021; Engle et al., 2020). Transition risk assessment maps corporate exposures to decarbonization scenarios—the Network for Greening the Financial System (NGFS) reference scenarios provide a standardized framework—and estimates the financial impact of policy changes, technology shifts, and demand substitution on firm valuations (Krueger et al., 2020; Bolton and Kacperczyk, 2021; NGF, ????). Engle et al. (2020) construct a hedging portfolio for climate change news, demonstrating that climate risk is already partially priced in equity markets, while Krueger et al. (2020) survey institutional investors and find that 70% consider climate risk in their investment process—a figure that has almost certainly increased since 2020 given subsequent regulatory mandates.

1.6.2 Digital Finance as Sustainability Enabler

Green bond markets demonstrate how digital finance infrastructure can accelerate sustainable capital allocation. Global green bond issuance exceeded USD 500 billion in 2023 and is projected to surpass USD 600 billion in 2024, making green bonds the fastest-growing segment of fixed income markets (CBI, ????). Flammer (2021) provides causal evidence that corporate green bonds improve environmental performance: firms that issue green bonds subsequently reduce CO_2 emissions relative to matched control firms, and the effect is concentrated among bonds with third-party verification—suggesting that the commitment mechanism, not merely the labeling, drives the outcome. Smart contract infrastructure can strengthen this mechanism further. Programmable use-of-proceeds restrictions encoded in bond covenants enable automated verification: funds released from a green bond escrow only when IoT sensors confirm that the financed project meets predefined environmental thresholds. This shifts verification from periodic reporting to continuous monitoring, addressing the "impact washing" problem in which green bond proceeds are fungible with general corporate funding.

Carbon markets exemplify both the promise and the peril of tokenization for sustainability. Tokenized carbon credits—Toucan Protocol bridged over 20 million tonnes of carbon offsets onto blockchain infrastructure before regulators intervened—promised to increase liquidity, reduce transaction costs, and democratize access to carbon markets (Tou, ????). In practice, the results were mixed. Blockchain infrastructure solved the double-counting problem that plagues voluntary carbon markets by providing immutable provenance tracking: each credit's vintage, methodology, and retirement status is recorded on-chain. But tokenization did not solve the deeper integrity challenges. Additionality—whether the emission reduction would have occurred without the credit—remains a methodological judgment that blockchain cannot automate. Wes (????) find that a substantial fraction of tokenized credits originated from projects with questionable additionality, meaning that blockchain's transparency made the quality problem more visible without resolving it. KlimaDAO's attempt to create a price floor for carbon credits through token-based demand generated temporary price spikes but collapsed when speculative demand withdrew, demonstrating that financialization of environmental assets introduces volatility that can undermine long-term market function.

Financial inclusion represents the social dimension of sustainability where digital finance has produced the most rigorous empirical evidence of positive impact. The World Bank Global Wor (????) reports that 1.4 billion adults worldwide lack access to formal financial services—a figure that has declined substantially from 2.5 billion in 2011, with digital finance as the primary driver of the reduction. M-Pesa, the mobile money platform launched in Kenya in 2007, provides the most extensively studied case. Sur (????) estimate that access to M-Pesa increased per capita consumption by 2% and lifted approximately 194,000 households—2% of Kenyan households—out of poverty between 2008 and 2014. The effects were concentrated among female-headed households, where consumption gains reached 3.6%, suggesting that mobile money's impact operates partly through the reduction of gendered barriers to financial access. Jac (????) document that M-Pesa users are better able to smooth consumption in response to income shocks, with remittance flows increasing by 15% following adverse events—a risk-sharing improvement that formal insurance markets have failed to deliver at comparable cost. These are not small effects. They demonstrate that digital payment infrastructure, even at low technological sophistication, generates welfare gains that traditional banking systems could not replicate in contexts where physical branch networks are economically unviable (Arner et al., 2020; Frost, 2020).

Whether DeFi protocols contribute to sustainability is a harder question. ESG-screened liquidity pools—Aave's proposed "green pool" that restricts collateral to verified green bonds and sustainability-linked tokens—represent genuine architectural innovation: the programmable nature of DeFi permits the embedding of sustainability criteria directly into protocol logic, where compliance is enforced by smart contract rather than by monitoring. Impact-linked protocols that adjust interest rates based on verified social outcomes create financial incentives at the protocol layer rather than through ad hoc corporate social responsibility programs. The critical assessment is that these innovations remain experimental, with total value locked in sustainability-focused DeFi protocols comprising less than 0.1% of the DeFi ecosystem as of early 2026. The gap between the theoretical potential and the deployed reality is large, and the history of DeFi suggests caution: composability and yield optimization have consistently attracted more capital than impact measurement.

1.6.3 Digital Finance as Sustainability Threat

The environmental cost of proof-of-work consensus mechanisms is not speculative. The Cambridge Bitcoin Electricity Consumption Index estimated Bitcoin's annualized electricity consumption at approximately 100 TWh in 2023—comparable to the total electricity consumption of the Netherlands, or roughly 0.4% of global electricity production (deV, ????a; Auer et al., 2022; Saleh, 2021). At the carbon intensity of the average mining geography, this translates to approximately 50–65 million tonnes of CO_2 annually. Ethereum's transition from proof-of-work to proof-of-stake in September 2022 reduced its energy consumption by an estimated 99.95%, demonstrating that the energy problem is consensus-mechanism-specific rather than blockchain-inherent (Saleh, 2021). Bitcoin's architecture precludes a similar transition without a governance event that its decentralized structure makes implausible. The energy cost is ongoing and, under current network economics, scales with Bitcoin's price: higher prices attract more mining hardware, which increases energy consumption until marginal mining revenue equals marginal electricity cost.

Electronic waste compounds the environmental liability. Application-specific integrated circuits (ASICs) designed for Bitcoin mining have useful lives of 18–24 months before they become economically obsolete. De deV (????b) estimate that the Bitcoin network generates approximately 30,700 metric tonnes of electronic waste annually—comparable to the small IT equipment waste of a country like the Netherlands—with each transaction generating the equivalent of discarding two iPhone 13 devices. Unlike general-purpose computing hardware, mining ASICs have no secondary use case; their entire lifecycle, from silicon fabrication to landfill, serves a single application.

Algorithmic bias in financial services is a social sustainability threat with documented empirical evidence. Bar (????) analyze 2.2 million mortgage applications processed by fintech lenders and find that Black and Hispanic borrowers pay 5.3 and 2.0 basis points more in interest, respectively, than White borrowers with identical creditworthiness—a finding that persists after controlling for credit score, loan-to-value ratio, and property characteristics. The magnitude is smaller than the discrimination documented in face-to-face lending (approximately 7.9 basis points), but the result is consequential: algorithmic lending reduces discrimination without eliminating it. Fus (????) demonstrate that machine learning models trained on historically biased data can reproduce discriminatory outcomes even when protected characteristics are excluded from the feature set, because correlated variables (zip code, education, employer) serve as proxies. The tradeoff between fairness metrics—demographic parity (equal approval rates across groups), equalized odds (equal true positive and false positive rates), and calibration (equal predictive accuracy across groups)—is mathematically irreconcilable: Cho (????) proves that except in trivial cases, no classifier can simultaneously satisfy all three. This means that algorithmic fairness in credit scoring is a design choice, not a technical problem, and the choice of which fairness criterion to optimize has distributional consequences that data science cannot resolve through better models.

The Jevons paradox operates in digital finance with particular force. Efficiency gains from algorithmic trading, automated compliance, and digital payment processing reduce the marginal cost of financial transactions, which increases transaction volume, which increases aggregate resource consumption. High-frequency trading firms execute millions of orders per day, the vast majority of which are cancelled within milliseconds (Aquilina et al., 2022; Budish et al., 2015); the energy cost per useful transaction is far higher than the cost per transaction would suggest. Data center energy consumption for AI training and inference in financial applications is growing at 25–30% annually, driven by the adoption of large language models for document processing, customer service, and risk analysis (IEA, ????). The efficiency gains are real; the rebound effects are also real; and the net environmental impact depends on which grows faster.

1.6.4 Toward Sustainability by Design

The European Union's regulatory architecture increasingly mandates what the market has not delivered voluntarily. The EU Taxonomy defines science-based technical screening criteria for environmentally sustainable economic activities. The Sustainable Finance Disclosure Regulation (SFDR) requires financial market participants to disclose how they integrate sustainability risks. The Corporate Sustainability Reporting Directive (CSRD) extends mandatory sustainability reporting to approximately 50,000 firms. Together, these instruments create a disclosure infrastructure that, for the first time, generates standardized sustainability data at the scale that quantitative analysis requires (Krueger et al., 2020; EUT, ????).

The design principles that emerge from the preceding analysis are concrete. Energy efficiency must be a default architectural choice, not an afterthought: proof-of-stake consensus, model compression for financial AI, and carbon-aware compute scheduling should be baseline requirements, not competitive differentiators. Fairness auditing must be embedded in the model development lifecycle, with explicit documentation of which fairness criteria are optimized and which tradeoffs are accepted—the EU AI Act's conformity assessment requirements point in this direction (EUA, ????). Impact measurement must be integrated into protocol design: green bond smart contracts that report environmental outcomes in real time, lending protocols that track financial inclusion metrics by construction, carbon market infrastructure that verifies additionality on-chain.

Data science is the connective tissue. The T_{ESG} scoring framework from Section 1.2.2 provides the measurement layer; the NLP, satellite, and ML methods discussed in this section provide the verification layer; the regulatory mandates provide the enforcement layer. None of these layers is sufficient alone. The argument of this book is that their convergence—data science methods applied to digital finance infrastructure under sustainability-oriented regulation—creates the conditions for sustainable financial innovation that is evidence-based rather than aspirational. Whether that potential is realized depends on implementation choices that the remaining chapters of this volume examine in detail.

1.7 Research Frontier and Open Questions

The preceding sections have established digital finance as a convergence phenomenon: electronic trading, artificial intelligence, blockchain infrastructure, and sustainability measurement are not independent trajectories but increasingly entangled systems whose interactions generate both the most significant opportunities and the most difficult research problems. The open questions that matter most are those that live at the intersection of multiple technology domains, where the analytical frameworks of any single discipline prove insufficient. This section maps the research frontier along the five research areas of the MSCA DIGITAL project (Grant Agreement No. 101119635) and identifies the cross-cutting themes that connect them.

1.7.1 European Financial Data Space

The European Commission's Financial Data Access framework (EUF, ????), together with the Data Governance Act (EUD, ????b) and the broader European Data EUD (????a), envisions a common financial data space in which institutions share information under standardized terms while preserving competitive confidentiality and individual privacy. The research frontier here is not regulatory design but technical feasibility. Cross-border financial data interoperability requires reconciling heterogeneous data models, legal regimes, and reporting frequencies across 27 member states, and no existing standard—neither ISO 20022 for payments nor the Legal Entity Identifier system (GLE, ????)—covers the full scope of what a functional data space demands. The question of which interoperability architecture can achieve semantic consistency without requiring centralized schema governance remains open.

Privacy-preserving computation offers a path toward data sharing for systemic risk monitoring without exposing proprietary positions. Federated learning enables multiple institutions to train joint models on distributed data without centralizing raw inputs (Yan, ????; Lon, ????), while secure multi-party computation allows regulators to compute aggregate risk statistics over encrypted bank-level data (Arc, ????). Fully homomorphic encryption extends this further by permitting arbitrary computation on ciphertexts (Arm, ????). Each of these approaches carries distinct tradeoffs in computational overhead, communication cost, and threat model assumptions, and their performance under realistic financial data conditions—irregular time series, fat-tailed distributions, high-dimensional cross-sectional panels—is not well characterized. Gol (????) document that alternative data already reshapes financial forecasting, but the quality benchmarks that would allow researchers to distinguish signal from noise in non-traditional datasets do not exist. Des (????) show that the forecasting value of alternative data degrades at longer horizons, raising the question of whether synthetic data generation methods can preserve the distributional properties that matter for financial applications—fat tails, volatility clustering, cross-asset dependence—while providing the volume that machine learning models require.

1.7.2 Artificial Intelligence in Financial Markets

The empirical asset pricing literature has established that machine learning models outperform linear factor models in cross-sectional return prediction (Gu et al., 2020; Feng et al., 2020), but this finding raises more questions than it resolves. Gu et al. (2020) demonstrate that neural networks and random forests achieve significant out-of-sample R^2 improvements, yet the economic interpretation of their predictions remains opaque. The critical open question is whether these models capture genuine causal structure in asset prices or exploit correlational patterns that are unstable under regime change. Chen et al. (2024) extend deep learning to conditional asset pricing and document substantial time-variation in factor loadings, but the methods available for causal inference in high-dimensional, non-stationary financial data are immature relative to the complexity of the problem.

Backtesting machine learning strategies under structural breaks presents a distinct methodological challenge. de Prado (2018) catalogues the overfitting pathologies that arise when researchers optimize strategies on historical data that includes regime transitions, and Har (????a) demonstrate that the standard statistical thresholds for evaluating trading signals are far too permissive given the scale of the multiple testing problem in empirical finance. Whether transformer architectures—which have demonstrated strong performance on sequential data in natural language processing—can adapt to the irregular, event-driven, and regime-switching nature of financial time series is an active area of investigation. Jiang et al. (2023) show that convolutional networks applied to price chart images extract predictive information that traditional time-series features miss, suggesting that the representation of financial data matters as much as the model architecture.

The rapid adoption of large language models in finance (Wu et al., 2023; Lopez-Lira and Tang, 2024; Osterrieder, 2023) introduces a new category of open problems. Lopez-Lira and Tang (2024) find that ChatGPT-based sentiment scores predict next-day stock returns, while Cao et al. (2023) document that firms have already begun adapting their disclosure language in response to algorithmic text analysis. Foundation models trained on general corpora lack the domain-specific calibration that financial risk management requires, and the evaluation frameworks for assessing when these models produce confident but incorrect financial reasoning do not yet exist. The question of how to deploy AI agents in trading and portfolio management with enforceable safety constraints—position limits, drawdown triggers, regulatory boundaries—that cannot be circumvented through prompt engineering or reward hacking is perhaps the most consequential near-term research problem in financial AI.

1.7.3 Explainable and Fair Artificial Intelligence

The EU AI Act (EUA, ????) classifies credit scoring and insurance pricing as high-risk applications subject to transparency, human oversight, and conformity assessment requirements. Which explainability methods satisfy these requirements in practice is not settled. Post-hoc explanation techniques—SHAP values, LIME, attention visualization—provide local interpretability but do not guarantee that the explanations are faithful to the model's actual decision process. The financial domain demands a form of interpretability that connects model outputs to the economic quantities that regulators and consumers understand: creditworthiness drivers, risk factor exposures, pricing components. Whether domain-specific interpretability methods that meet this standard can be developed without sacrificing the predictive gains documented by Gu et al. (2021) is an open empirical question, not a foregone conclusion.

Algorithmic fairness in financial services confronts Cho's ???? impossibility result: demographic parity, equalized odds, and calibration cannot be simultaneously satisfied except in degenerate cases. The research question is not how to resolve the impossibility but how to operationalize the choice. Bar (????) show that fintech lending reduces but does not eliminate racial disparities, and Fus (????) demonstrate that proxy discrimination persists even when protected characteristics are excluded from model inputs. Developing bias auditing procedures that are both computationally tractable and regulatorily defensible—capable of detecting disparate impact across intersectional subgroups in high-dimensional feature spaces—remains an unsolved problem at the intersection of computer science, law, and economics.

1.7.4 Blockchain and Decentralized Finance

The blockchain scalability trilemma—the conjecture that no system can simultaneously achieve decentralization, security, and throughput—has driven a proliferation of Layer 2 solutions: optimistic rollups, zero-knowledge rollups, and sidechains (But, ????). Whether these represent permanent architectural features or transitional patches awaiting a more fundamental solution shapes the entire trajectory of blockchain-based financial infrastructure. Cross-chain interoperability is equally unresolved: bridge protocols have suffered billions of dollars in exploits, and the alternative approaches—shared sequencers, cross-chain message-passing protocols—introduce their own trust assumptions that have not been formally verified at scale (Werner et al., 2022; Aramonte et al., 2021).

In decentralized finance, the open research agenda extends well beyond protocol design. Schär (2021) provides the canonical taxonomy of DeFi protocols, but the governance mechanisms that would make these systems accountable to their users—beyond plutocratic token voting—are underdeveloped (Zetzsche et al., 2020). Maximal extractable value (MEV), documented by Daian et al. (2020), creates systematic wealth transfers from ordinary users to sophisticated actors; whether MEV can be mitigated without reintroducing the centralized intermediaries that DeFi was designed to eliminate is a deep structural question. Stablecoin design, following the Terra/Luna collapse (Uhlig, 2022; Clements, 2021), requires mechanism design frameworks that account for reflexive dynamics under stress: the conditions under which algorithmic stabilization fails are precisely the conditions under which stability matters most.

1.7.5 Sustainability of Digital Finance

Section 1.6 documented that ESG rating divergence ($\rho \approx 0.54$ across providers) constitutes a first-order obstacle to evidence-based sustainable finance (Berg et al., 2022). The research frontier is not standardization for its own sake but whether standardization can be achieved without eliminating the variation that reflects genuine methodological disagreement about what sustainability means. The EU EUT (????) and CSRD provide regulatory convergence pressure, but their technical screening criteria require input data—Scope 3 emissions, biodiversity impact metrics, social governance indicators—that is measured with uncertainty far exceeding what financial risk models are calibrated to handle.

Greenwashing detection at scale requires methods that control false positive rates at acceptable levels for regulatory enforcement. Bin (????) and Luc (????) demonstrate that NLP-based approaches can identify discrepancies between corporate climate commitments and implemented actions, but the transition from academic proof-of-concept to regulatory-grade impact methodology—with defined error rates, audit trails, and legal defensibility—has not been made. Bolton and Kacperczyk (2021) and Engle et al. (2020) show that climate risk is at least partially priced in equity markets, yet integrated assessment models that quantitatively link digital finance interventions (green bond issuance, carbon credit tokenization, ESG-screened lending) to measurable environmental outcomes remain scarce. The Jevons paradox documented in Section 1.6.3—efficiency gains that increase aggregate resource consumption—demands longitudinal studies that track net impact rather than marginal improvements, and the data infrastructure for such studies does not yet exist at the resolution required (IEA, ????; Giglio et al., 2021).

1.7.6 Cross-Cutting Themes

The most important feature of these open questions is that none of them can be resolved within a single discipline. Privacy-preserving financial data sharing requires cryptography, financial econometrics, and regulatory law. Algorithmic fairness in credit scoring spans computer science, economics, and civil rights law. Sustainable finance measurement demands environmental science, accounting, and statistical methodology. The research frontier in digital finance is, in this precise sense, an interdisciplinary frontier, and progress depends on building institutions and incentive structures that support genuinely collaborative work across disciplinary boundaries (Goldstein et al., 2019; Allen et al., 2021; Boot et al., 2021).

Reproducibility infrastructure is a precondition for progress on all five research areas. The Quantlet platform (Har, ????b) and the Quantinar collaborative research environment (Har, ????c) provide mechanisms for sharing code, data pipelines, and computational workflows in a form that enables independent verification and extension. The empirical findings reviewed in this chapter—from the asset pricing results of Gu et al. (2020) to the ESG rating divergence documented by Berg et al. (2022)—are only as robust as the ability of other researchers to replicate and stress-test them. Hou (????) demonstrate what happens when replication is taken seriously: a substantial fraction of published anomalies fail to survive independent verification.

The subsequent chapters of this volume address specific segments of this frontier. Chapter 2 develops the data science and financial risk assessment methods that underpin the European Financial Data Space agenda. Chapter 3 examines innovations in fintech and digital markets through the lens of the AI and blockchain research questions identified here. Chapter 4 confronts the policy implications—regulatory design, fairness requirements, sustainability mandates—that shape which technological possibilities become deployed realities. The industry and policy insight chapters ground these academic questions in practitioner experience and institutional context. The argument of this book is that the convergence documented throughout this chapter is not merely a descriptive observation but a research program: the most consequential advances in digital finance will come from work that operates simultaneously across the technology, methodology, and governance dimensions that any single perspective cannot span.

1.8 Conclusion and Outlook

This chapter has surveyed digital finance as a convergence phenomenon—a set of technological, regulatory, and analytical trajectories that are no longer separable and whose interactions define the field's most consequential opportunities and risks. The analysis yields several findings that warrant explicit statement.

First, digital finance is not a collection of independent technologies but an increasingly unified programmable infrastructure. Algorithmic trading strategies now execute on decentralized platforms, AI models score the creditworthiness of DeFi borrowers, and blockchain settlement requires machine learning for compliance verification. The three-dimensional taxonomy developed in this chapter—organizing innovations along functional, technological, and disruptive axes—captures this entanglement and reveals that the most significant developments occur at domain boundaries rather than within any single technology silo (Boot et al., 2021; Goldstein et al., 2019).

Second, the technology foundations reviewed here demonstrate that each of the five core domains—electronic trading, blockchain infrastructure, artificial intelligence, DeFi and tokenization, and digital banking—has matured beyond proof-of-concept into production deployment. Spot Bitcoin ETFs attracted record-breaking institutional inflows (Bla, ????; Dowling and Lucey, 2024). Machine learning models achieve statistically significant out-of-sample improvements in asset pricing over linear factor models (Gu et al., 2020). Large language models have moved from research curiosities to operational tools for regulatory reporting, risk narrative generation, and sentiment extraction (Wu et al., 2023; Lopez-Lira and Tang, 2024; Osterrieder, 2023). These are no longer speculative capabilities; they are deployed realities whose governance demands immediate attention.

Third, data infrastructure constitutes the binding constraint on whether convergence materializes as practice rather than aspiration. The European Financial Data Space envisions cross-border interoperability under standardized terms, but reconciling heterogeneous data models, legal regimes, and reporting frequencies across 27 member states remains an unsolved technical and institutional problem. The gap between the data that digital finance technologies require and the data that existing infrastructure provides shapes the trajectory of every domain reviewed in this chapter.

Fourth, ESG measurement divergence is not a technical inconvenience but a first-order obstacle to evidence-based sustainable finance. The average pairwise correlation of $\rho \approx 0.54$ across major ESG rating providers (Berg et al., 2022) reflects fundamental disagreement about what sustainability is, not merely how to quantify it. Data science offers partial resolution through NLP-based analysis of corporate disclosures (Bin, ????), satellite-based independent verification of environmental claims (Giglio et al., 2021; Bolton and Kacperczyk, 2021), and machine learning approaches to composite score construction—but aggregation without addressing scope divergence risks producing estimates that are precise without being accurate.

Fifth, digital finance operates simultaneously as enabler and threat to sustainability objectives. Green bond infrastructure, carbon credit tokenization, and mobile money platforms have delivered measurable positive outcomes: Flammer (2021) documents causal emission reductions from corporate green bonds; Sur (????) estimate that mobile money lifted 194,000 Kenyan households out of poverty. Yet proof-of-work consensus mechanisms consume electricity at national scale (deV, ????a), algorithmic lending reduces but does not eliminate racial disparities in credit pricing (Bar, ????; Fus, ????), and the Jevons paradox ensures that efficiency gains from automation increase aggregate resource consumption when transaction costs fall (Aquilina et al., 2022; IEA, ????). Acknowledging this duality honestly is a precondition for designing systems that produce net positive outcomes.

Sixth, sustainability-by-design must become an architectural default rather than a retrospective overlay. The EU regulatory stack—the Taxonomy, SFDR, CSRD, MiCA, DORA, and the AI Act—increasingly mandates what the market has not delivered voluntarily. The design principles are concrete: proof-of-stake consensus as baseline, fairness auditing embedded in the model development lifecycle, impact measurement integrated into protocol logic. Data science provides the connective tissue—the measurement, verification, and enforcement layers—without which sustainability in finance remains aspirational rather than actionable (Krueger et al., 2020; EUA, ????; EUT, ????).

Seventh, the failures examined in this chapter—Terra/Luna, FTX, equity market flash crashes—are not aberrations but structural consequences of composability without adequate governance. The same interconnectedness that enables rapid innovation creates fragility: concentrated algorithmic strategies, opaque smart contracts, and reflexive feedback loops amplify tail risk in ways that traditional financial risk models were not designed to capture (Uhlig, 2022; Clements, 2021; Kirilenko et al., 2017).

Eighth, the research frontier mapped in this chapter is irreducibly interdisciplinary. Privacy-preserving financial data sharing requires cryptography, financial econometrics, and regulatory law. Algorithmic fairness in credit scoring spans computer science, economics, and civil rights law. Sustainable finance measurement demands environmental science, accounting, and statistical methodology. Progress depends on building institutions and incentive structures that support genuinely collaborative work across disciplinary boundaries (Allen et al., 2021).

Looking ahead three to five years, several developments will reshape the landscape surveyed here. The Markets in Crypto-Assets Regulation is now in full application (MiC, ????; ESM, ????), and its effects on European market structure are beginning to materialize: licensing requirements will consolidate the fragmented crypto-asset service provider landscape, stablecoin reserve mandates will constrain the issuance models that proved catastrophic in the Terra/Luna episode, and the transparency obligations will generate structured data that enables the kind of quantitative analysis this field has lacked. Whether MiCA achieves regulatory coherence or drives activity to less regulated jurisdictions depends on the calibration of its implementing measures and on the degree to which other major economies—particularly the United States and the United Kingdom—converge on compatible frameworks. Central bank digital currencies will advance from pilot programs to deployment decisions. The European Central Bank's Digital Euro project has moved through its investigation and preparation phases; the critical design choices—offline functionality, privacy architecture, holding limits, and interoperability with existing payment infrastructure (ECB, ????; Auer et al., 2024)—will determine whether the Digital Euro becomes a meaningful addition to the European payments landscape or a solution in search of a problem. International CBDC interoperability, pursued through projects such as mBridge, raises deeper questions about monetary sovereignty, cross-border capital flows, and the future architecture of correspondent banking (Kosse and Mattei, 2023; Agur et al., 2022).

The EU AI Act's implementation timeline will test whether risk-based regulation of artificial intelligence can keep pace with the technology it governs. For financial services, the high-risk classification of credit scoring and insurance pricing imposes transparency, human oversight, and conformity assessment requirements whose practical interpretation remains unsettled (EUA, ????). The deployment of AI agents in trading, portfolio management, and regulatory compliance introduces a category of operational risk—autonomous systems making consequential financial decisions under conditions their training data did not anticipate—for which existing supervisory frameworks offer limited guidance.

Real-world asset tokenization will likely advance selectively. Government securities and high-grade corporate bonds—standardized, liquid, and well-understood from a legal perspective—represent the asset classes where tokenization offers the clearest efficiency gains through reduced settlement times and programmable compliance. Less standardized asset classes—real estate, private equity, infrastructure—face legal and valuation complexities that blockchain infrastructure does not resolve and may complicate (Nair et al., 2024). The sustainability dimension of tokenization—whether programmable use-of-proceeds restrictions and automated impact verification can strengthen green bond markets—will depend on whether the infrastructure develops at the intersection of financial engineering and environmental measurement rather than as a purely financial innovation.

Several uncertainties cut across all of these trajectories. Geopolitical fragmentation threatens the interoperability on which a global digital finance infrastructure depends: divergent regulatory approaches between the EU, the US, and China risk creating isolated digital finance zones whose incompatibility imposes real economic costs. Quantum computing advances, while not yet an immediate operational threat, impose a planning horizon for cryptographic migration that responsible infrastructure design must incorporate. The pace at which regulatory frameworks adapt to technological change—consistently slower than the technology itself—determines whether regulation shapes innovation toward socially desirable outcomes or merely constrains the form it takes. And data science, the central methodological thread of this volume, must itself evolve: the methods that bridge innovation and sustainability today—NLP-based greenwashing detection, climate risk modeling, federated learning for privacy-preserving data sharing—are necessary but not sufficient for the next generation of challenges that quantum-resistant cryptography, autonomous AI agents, and integrated sustainability measurement will present.

This foundational chapter has established the conceptual, technological, and regulatory context within which the remainder of this volume operates. The convergence thesis articulated here—that digital finance technologies, data science methodologies, and sustainability imperatives are becoming inseparable—frames the questions that subsequent treatments address. How should financial risk assessment adapt when the data generating processes are themselves algorithmic? What institutional innovations in fintech and digital markets will prove durable, and which represent transient arbitrage of regulatory gaps? How should policy frameworks balance the promotion of innovation against the containment of systemic risk? What does practitioner experience reveal about the gap between academic models and deployed reality? And what research frontiers, identified in the preceding section, offer the highest return on interdisciplinary investment? These questions are not rhetorical. They define the research program that this volume, and the MSCA DIGITAL doctoral network from which it emerges, is designed to advance.

References

(????)
(????)
(????)
(????a) Tech. rep.
(????) Tech. rep.
(????)
(????)
(????)
(????)
(????) Tech. rep.
(????) Tech. rep.
(????) Tech. rep.
(????) Tech. rep.
(????)
(????a) Tech. rep.
(????) Tech. rep.
(????b) Tech. rep.
(????)
(????)
(????) Tech. rep.
(????b) Tech. rep.
(????) Tech. rep.
(????)
(????) Tech. rep.
(????)
(????)
(????a)
(????)
(????a) Tech. rep.
(????) Tech. rep.
(????b) Tech. rep.
(????)
(????)
(????)
(????)
(????)
(????b)
(????c)
(????) Tech. rep.
(????)

(????)
(????)
(????)
(????) Tech. rep.
(????) Tech. rep.
(????) Tech. rep.
(????)
(????) Tech. rep.
(????)
(????)
(????a)
(????b)
(????)
(????)
(????) Tech. rep.
(????) Tech. rep.

Agur I, Ari A, Dell'Ariccia G (2022) Designing central bank digital currencies. Journal of Monetary Economics 125:62–79, DOI 10.1016/j.jmoneco.2021.05.002

Allen F, Gu X, Jagtiani J (2021) A survey of fintech research and policy discussion. Review of Corporate Finance 1(3–4):259–339, DOI 10.1561/114.00000007

Allen F, Gu X, Jagtiani J (2022) Fintech, cryptocurrencies, and CBDC: Financial structural transformation in China. Journal of International Money and Finance 124:102625, DOI 10.1016/j.jimonfin.2022.102625

Almgren R, Chriss N (2001) Optimal execution of portfolio transactions. Journal of Risk 3(2):5–39, DOI 10.21314/JOR.2001.041

Anagnostopoulos I (2018) Fintech and regtech: Impact on regulators and banks. Journal of Economics and Business 100:7–25, DOI 10.1016/j.jeconbus.2018.07.003

Angeris G, Kao HT, Chiang R, Noyes C, Chitra T (2021) An analysis of Uniswap markets. Cryptoeconomic Systems 1(1)

Aquilina M, Budish E, O'Neill P (2022) Quantifying the high-frequency trading "arms race". Quarterly Journal of Economics 137(1):493–564, DOI 10.1093/qje/qjab032

Aramonte S, Huang W, Schrimpf A (2021) DeFi risks and the decentralisation illusion. BIS Quarterly Review pp 21–36

Aramonte S, Schrimpf A, Shin HS (2023) Non-bank financial intermediaries and financial stability. BIS Working Papers (972)

Arner DW, Barberis J, Buckley RP (2017) Fintech, regtech, and the reconceptualization of financial regulation. Northwestern Journal of International Law and Business 37(3):371–413

Arner DW, Buckley RP, Zetzsche DA, Veidt R (2020) Sustainability, fintech and financial inclusion. European Business Organization Law Review 21:7–35, DOI 10.1007/s40804-020-00183-y

Auer R, Cornelli G, Frost J (2020) Rise of the central bank digital currencies: Drivers, approaches and technologies. BIS Working Papers (880)

Auer R, Farag M, Lewrick U, Orazem L, Zoss M (2022) Banking in the shadow of Bitcoin? the institutional adoption of cryptocurrencies. BIS Working Papers (1013)

Auer R, Frost J, Lee M, Martin A, Narula N (2024) Why are central banks interested in CBDC? BIS Working Papers

Avramov D, Cheng S, Lioui A, Tarelli A (2022) Sustainable investing with ESG rating uncertainty. Journal of Financial Economics 145(2):642–664, DOI 10.1016/j.jfineco.2021.09.009

Baron M, Brogaard J, Hagströmer B, Kirilenko A (2019) Risk and return in high-frequency trading. Journal of Financial and Quantitative Analysis 54(3):993–1024, DOI 10.1017/S0022109018001096

Bartram SM, Branke J, Motahari M (2020) Artificial intelligence in asset management. CFA Institute Research Foundation DOI 10.2139/ssrn.3692805

Bazot G (2018) Financial consumption and the cost of finance: Measuring financial efficiency in Europe (1950–2007). Journal of the European Economic Association 16(1):123–160, DOI 10.1093/jeea/jvx008

Berg F, Kölbel JF, Rigobon R (2022) Aggregate confusion: The divergence of ESG ratings. Review of Finance 26(6):1315–1344, DOI 10.1093/rof/rfac033

Berg T, Burg V, Gombović A, Puri M (2020) On the rise of fintechs: Credit scoring using digital footprints. Review of Financial Studies 33(7):2845–2897, DOI 10.1093/rfs/hhz099

Biais B, Bisiere C, Bouvard M, Casamatta C (2019) The blockchain folk theorem. Review of Financial Studies 32(5):1662–1715, DOI 10.1093/rfs/hhy095

Bolton P, Kacperczyk M (2021) Do investors care about carbon risk? Journal of Financial Economics 142(2):517–549, DOI 10.1016/j.jfineco.2021.05.008

Boot AWA, Hoffmann P, Laeven L, Ratnovski L (2021) Fintech: What's old, what's new? Journal of Financial Stability 53:100836, DOI 10.1016/j.jfs.2020.100836

Bouchaud JP, Farmer JD, Lillo F (2009) How markets slowly digest changes in supply and demand. Handbook of Financial Markets: Dynamics and Evolution pp 57–160, DOI 10.1016/B978-012374258-2.50006-3

Broeders D, Prenio J (2018) Innovative technology in financial supervision (SupTech)—the experience of early users. FSI Insights on Policy Implementation (9)

Brogaard J, Hendershott T, Riordan R (2014) High-frequency trading and price discovery. Review of Financial Studies 27(8):2267–2306, DOI 10.1093/rfs/hhu032

Brunnermeier MK (2009) Deciphering the liquidity and credit crunch 2007–2008. The Journal of Economic Perspectives 23:77–100, DOI 10.1257/jep.23.1.77

Brunnermeier MK, James H, Landau JP (2019) The digitalization of money. NBER Working Paper (26300), DOI 10.3386/w26300

Buchak G, Matvos G, Piskorski T, Seru A (2018) Fintech, regulatory arbitrage, and the rise of shadow banks. Journal of Financial Economics 130(3):453–483, DOI 10.1016/j.jfineco.2018.03.011

Buckley RP, Arner DW, Zetzsche DA, Selga E (2020) Techrisk: Fintech and regulatory risk. European Banking Institute Working Paper (60)

Budish E, Cramton P, Shim J (2015) The high-frequency trading arms race: Frequent batch auctions as a market design response. Quarterly Journal of Economics 130(4):1547–1621, DOI 10.1093/qje/qjv027

Buehler H, Gonon L, Teichmann J, Wood B (2019) Deep hedging. Quantitative Finance 19(8):1271–1291, DOI 10.1080/14697688.2019.1571683

Campbell JY, Lo AW, MacKinlay AC (2012) The econometrics of financial markets. DOI 10.2307/j.ctt7skm5

Cao S, Jiang W, Yang B, Zhang AL (2023) How to talk when a machine is listening: Corporate disclosure in the age of AI. Review of Financial Studies 36(9):3603–3642, DOI 10.1093/rfs/hhad021

Capponi A, Jia R (2021) The adoption of blockchain-based decentralized exchanges. Working Paper DOI 10.2139/ssrn.3805095

Cartea Á, Jaimungal S, Penalva J (2015) Algorithmic and high-frequency trading

Chen L, Pelger M, Zhu J (2024) Deep learning in asset pricing. Management Science 70(2):714–750, DOI 10.1287/mnsc.2023.4695

Chordia T, Roll R, Subrahmanyam A (2000) Commonality in liquidity. Journal of Financial Economics 56:3–28, DOI 10.1016/s0304-405x(99)00057-4

Claessens S, Frost J, Turner G, Zhu F (2018) Fintech credit markets around the world: Size, drivers and policy issues. BIS Quarterly Review pp 29–49

Clements R (2021) Built to fail: The inherent fragility of algorithmic stablecoins. Wake Forest Law Review Online 11:131–171

Cong LW, He Z (2019) Blockchain disruption and smart contracts. Review of Financial Studies 32(5):1754–1797, DOI 10.1093/rfs/hhz007

Cont R (2011) Statistical modeling of high-frequency financial data. IEEE Signal Processing Magazine 28(5):16–25, DOI 10.1109/MSP.2011.941548

Cornelli G, Frost J, Gambacorta L, Rau R, Wardrop R, Ziegler T (2023) Fintech and big tech credit: Drivers of the growth of digital lending. Journal of Banking & Finance 148:106742, DOI 10.1016/j.jbankfin.2022.106742

Daian P, Goldfeder S, Kell T, Li Y, Zhao X, Bentov I, Breidenbach L, Juels A (2020) Flash boys 2.0: Frontrunning in decentralized exchanges, miner extractable value, and consensus instability. In: 2020 IEEE Symposium on Security and Privacy, pp 910–927, DOI 10.1109/SP40000.2020.00040

Demirgüç-Kunt A, Klapper L, Singer D, Ansar S, Hess J (2018) The global findex database 2017: Measuring financial inclusion and the fintech revolution. DOI 10.1596/978-1-4648-1259-0

Dixon MF, Halperin I, Bilokon P (2020) Machine Learning in Finance: From Theory to Practice. Springer, DOI 10.1007/978-3-030-41068-1

Dowling M, Lucey B (2024) Spot Bitcoin ETFs and crypto market structure. Finance Research Letters 61:105022, DOI 10.1016/j.frl.2024.105022

Drempetic S, Klein C, Zwergel B (2019) The influence of firm size on the esg score: Corporate sustainability ratings under review. Journal of Business Ethics 167:333–360, DOI 10.1007/s10551-019-04164-1

Easley D, O'Hara M (2004) Information and the cost of capital. The Journal of Finance 59:1553–1583, DOI 10.1111/j.1540-6261.2004.00672.x

Engle RF, Giglio S, Kelly B, Lee H, Stroebel J (2020) Hedging climate change news. Review of Financial Studies 33(3):1184–1216, DOI 10.1093/rfs/hhz072

Fama EF, French KR (2004) The capital asset pricing model: Theory and evidence. The Journal of Economic Perspectives 18:25–46, DOI 10.1257/0895330042162430

Feng G, Giglio S, Xiu D (2020) Taming the factor zoo: A test of new factors. Journal of Finance 75(3):1327–1370, DOI 10.1111/jofi.12883

Fernández-Villaverde R, Sanches D, Schilling L, Uhlig H (2021) Central bank digital currency: Central banking for all? Review of Economic Dynamics 41:225–242, DOI 10.1016/j.red.2020.12.004

Flammer C (2021) Corporate green bonds. Journal of Financial Economics 142(2):499–516, DOI 10.1016/j.jfineco.2021.01.010

Foucault T, Hombert J, Roşu I (2016) News trading and speed. Journal of Finance 71(1):335–382, DOI 10.1111/jofi.12302

Frost J (2020) The economic forces driving fintech adoption across countries. BIS Working Papers (838)

Fuster A, Plosser M, Schnabl P, Vickery J (2019) The role of technology in mortgage lending. Review of Financial Studies 32(5):1854–1899, DOI 10.1093/rfs/hhz018

Gentzkow M, Kelly B, Taddy M (2019) Text as data. Journal of Economic Literature 57(3):535–574, DOI 10.1257/jel.20181020

Giglio S, Kelly B, Stroebel J (2021) Climate finance. Annual Review of Financial Economics 13:15–36, DOI 10.1146/annurev-financial-102620-103311

Goldstein I, Jiang W, Karolyi GA (2019) To fintech and beyond. Review of Financial Studies 32(5):1647–1661, DOI 10.1093/rfs/hhz025

Goldstein I, Jiang W, Yu J (2023) The cryptocurrency market: A study of market microstructure. Working Paper

Gomber P, Koch JA, Siering M (2017) Digital finance and fintech: current research and future research directions. Journal of Business Economics 87:537–580, DOI 10.1007/s11573-017-0852-x

Griffin JM, Shams A (2020) Is Bitcoin really untethered? Journal of Finance 75(4):1913–1964, DOI 10.1111/jofi.12903

Gu S, Kelly B, Xiu D (2020) Empirical asset pricing via machine learning. Review of Financial Studies 33(5):2223–2273, DOI 10.1093/rfs/hhaa009

Gu S, Kelly B, Xiu D (2021) Autoencoder asset pricing models. Journal of Econometrics 222(1):429–450, DOI 10.1016/j.jeconom.2020.07.009

Gudgeon L, Werner S, Perez D, Knottenbelt WJ (2020) DeFi protocols for loanable funds: Interest rates, liquidity and market efficiency. In: Proceedings of the 2nd ACM Conference on Advances in Financial Technologies, pp 92–112, DOI 10.1145/3419614.3423254

Guéant O (2016) The financial mathematics of market liquidity: From optimal execution to market making

Harvey CR, Ramachandran A, Santoro J (2021) DeFi and the Future of Finance. Wiley

Hendershott T, Jones CM, Menkveld AJ (2011) Does algorithmic trading improve liquidity? Journal of Finance 66(1):1–33, DOI 10.1111/j.1540-6261.2010.01624.x

Israel R, Kelly B, Moskowitz TJ (2020) Can machines "learn" finance? Journal of Investment Management 18(2):23–36

Jagtiani J, Lemieux C (2018) Do fintech lenders penetrate areas that are underserved by traditional banks? Journal of Economics and Business 100:43–54, DOI 10.1016/j.jeconbus.2018.03.001

Jiang J, Kelly B, Xiu D (2023) (re-)imag(in)ing price trends. Journal of Finance 78(6):3193–3249, DOI 10.1111/jofi.13268

Keister T, Sanches D (2023) Should central banks issue digital currency? Review of Economic Studies 90(1):404–431, DOI 10.1093/restud/rdac017

Kirilenko A, Kyle AS, Samadi M, Tuzun T (2017) The flash crash: High-frequency trading in an electronic market. Journal of Finance 72(3):967–998, DOI 10.1111/jofi.12498

Kosse A, Mattei I (2023) Gaining momentum: Results of the 2022 BIS survey on central bank digital currencies. BIS Papers (136)

Kozak S, Nagel S, Santosh S (2020) Shrinking the cross-section. Journal of Financial Economics 135(2):271–292, DOI 10.1016/j.jfineco.2019.06.008

Krueger P, Sautner Z, Starks LT (2020) The importance of climate risks for institutional investors. Review of Financial Studies 33(3):1067–1111, DOI 10.1093/rfs/hhz137

Kyle AS (1985) Continuous auctions and insider trading. Econometrica 53(6):1315–1335, DOI 10.2307/1913210

Lee I, Shin YJ (2017) Fintech: Ecosystem, business models, investment decisions, and challenges. Business Horizons 61:35–46, DOI 10.1016/j.bushor.2017.09.003

Lehar A, Parlour CA (2021) Decentralized exchanges. Working Paper DOI 10.2139/ssrn.3905316

Li X, Jiang P, Chen T, Luo X, Wen Q (2017) A survey on the security of blockchain systems. Future Generation Computer Systems 107:841–853, DOI 10.1016/j.future.2017.08.020

Li Y, Mayer S, Scheidegger S (2024) Cryptocurrency regulation and market quality. Journal of Financial Economics Forthcoming

Lopez-Lira A, Tang Y (2024) Can ChatGPT forecast stock price movements? return predictability and large language models. Journal of Finance DOI 10.2139/ssrn.4412788, forthcoming

Loughran T, McDonald B (2011) When is a liability not a liability? textual analysis, dictionaries, and 10-ks. Journal of Finance 66(1):35–65, DOI 10.1111/j.1540-6261.2010.01625.x

Mehar MI, Shier CL, Giambattista A, Gong E, Fletcher G, Sanayhie R, Kim HM, Lasber M (2019) Understanding a revolutionary and flawed grand experiment in blockchain: The DAO attack. Journal of Cases on Information Technology 21(1):19–32, DOI 10.4018/JCIT.2019010102

Menkveld AJ (2013) High frequency trading and the new market makers. Journal of Financial Markets 16(4):712–740, DOI 10.1016/j.finmar.2013.06.006

Menkveld AJ (2016) The economics of high-frequency trading: Taking stock. Annual Review of Financial Economics 8:1–24, DOI 10.1146/annurev-financial-121415-033010

Nair M, Arazzi M, Aarons Z (2024) Real world asset tokenization: Bridging traditional finance and DeFi. Working Paper

Nie Y, Cao Y, Kawaguchi K (2024) Large language models for financial applications: A survey. Working Paper DOI 10.2139/ssrn.4723250

O'Hara M (2015) High frequency market microstructure. Journal of Financial Economics 116(2):257–270, DOI 10.1016/j.jfineco.2015.01.003

Osterrieder J (2023) Large language models and generative AI for finance. Working Paper

Osterrieder J, Lorenz J (2017) A statistical risk assessment of Bitcoin and its extreme tail behavior. Annals of Financial Economics 12(1):1750003, DOI 10.1142/S2010495217500038

Osterrieder J, Misheva BH, Mare C (2024) Digital finance: Reaching new frontiers. SSRN Working Paper

Park A (2021) The conceptual flaws of constant product automated market making. Working Paper DOI 10.2139/ssrn.3805750

Pástor L, Stambaugh RF, Taylor LA (2021) Sustainable investing in equilibrium. Journal of Financial Economics 142(2):550–571, DOI 10.1016/j.jfineco.2020.12.011

Pedersen LH, Fitzgibbons S, Pomorski L (2021) Responsible investing: The ESG-efficient frontier. Journal of Financial Economics 142(2):572–597, DOI 10.1016/j.jfineco.2020.11.001

Philippon T (2016) The fintech opportunity. NBER Working Paper (22476), DOI 10.3386/w22476

de Prado ML (2018) Advances in financial machine learning

Saleh F (2021) Blockchain without waste: Proof-of-stake. Review of Financial Studies 34(3):1156–1190, DOI 10.1093/rfs/hhaa075

Schär F (2021) Decentralized finance: On blockchain- and smart contract-based financial markets. Federal Reserve Bank of St Louis Review 103(2):153–174, DOI 10.20955/r.103.153-74

Sirignano J, Sadhwani A, Giesecke K (2023) Deep learning for mortgage risk. Journal of Financial Economics 147(3):750–783, DOI 10.1016/j.jfineco.2023.01.002

Stulz RM (2019) Fintech, bigtech, and the future of banks. Journal of Applied Corporate Finance 31(4):86–97, DOI 10.1111/jacf.12378

Thakor AV (2020) Fintech and banking: What do we know? Journal of Financial Intermediation 41:100833, DOI 10.1016/j.jfi.2019.100833

Uhlig H (2022) A Luna-tic stablecoin crash. NBER Working Paper (30256), DOI 10.3386/w30256

Vives X (2019) Competition and stability in modern banking: A post-crisis perspective. International Journal of Industrial Organization 64:55–69, DOI 10.1016/j.ijindorg.2018.08.011

Werner SM, Perez D, Gudgeon L, Klages-Mundt A, Harz D, Knottenbelt WJ (2022) SoK: Decentralized finance (DeFi). Proceedings of the 4th ACM Conference on Advances in Financial Technologies pp 1–16, DOI 10.1145/3558535.3559780

Wu S, Irsoy O, Lu S, Dabravolski V, Dredze M, Gehrmann S, Kambadur P, Rosenberg D, Mann G (2023) BloombergGPT: A large language model for finance. arXiv preprint arXiv:230317564

Yermack D (2017) Corporate governance and blockchains. Review of Finance 21(1):7–31, DOI 10.1093/rof/rfw074

Zetzsche DA, Arner DW, Buckley RP (2020) Decentralized finance. Journal of Financial Regulation 6(2):172–203, DOI 10.1093/jfr/fjaa010

Acknowledgements This project has received funding from the European Union's Horizon Europe research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 101119635 (DIGITAL – Digital Finance: Reaching New Frontiers).