



International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Methodological Review of Blockchain-Based Verification and AI-Assisted Document Validation for Government NOC Management

Abhishek Bacche¹, Ayush Vyavhare², and Pranali Balwante³

Department of Computer Engineering, Smt. Indira Gandhi College of Engineering, Navi Mumbai, India. Corresponding author: Abhishek Bacche

ABSTRACT :

If you want to run a hotel, factory, or commercial building, you first need a No Objection Certificate (NOC) or a license. People still check those documents mostly by hand, on paper, and it's hard to audit once approval is given. Several technologies tackle pieces of this problem: blockchain hash anchoring, smart contracts, optical character recognition (OCR), document-understanding models, forgery detection, and human-in-the-loop decision support. This paper reviews 50 sources to see how these techniques have been built and tested, and how they might fit together into a verification workflow for government compliance certificates. Instead of sorting the technologies by type, we look at the job each one does: tamper-evident recording, off-chain storage, validation before approval, forgery detection, officer accountability and lifecycle tracking. Our reading is that hash anchoring and AI-based extraction are each fairly mature. What is rare is a study that combines them with officer-authoritative approval and expiry and renewal management, or that tests the whole chain end to end. We call this gap methodological fragmentation. From it we derive requirements for the next generation of NOC verification systems: AI-assisted preliminary validation, officer-authoritative approval, SHA-256 fingerprints with on-chain approval metadata, off-chain document storage, role-based access control, and automated expiry and renewal versioning. Together, these give a traceable path from the literature to a system design.

Index Terms: blockchain, NOC verification, smart contracts, SHA-256, document hashing, OCR, document validation, forgery detection, e-governance, role-based access control, compliance management.

Literature Search and Source Selection Methodology

This paper is a focused narrative review. It looks at research and standards on blockchain record integrity, document AI, forgery detection, credential verification, access control and compliance lifecycle management. It is not a registered systematic review or a meta-analysis, and we do not present it as one.

Search approach. We found sources through targeted searches of publisher pages, conference proceedings, standards bodies and research repositories. Search terms were combinations of phrases such as "blockchain document verification," "government e-governance blockchain," "credential verification," "OCR document understanding," "document forgery detection," "SHA-256," "role-based access control" and "AI human oversight." We also followed the reference lists of relevant papers and standards to reach foundational work.

Selection criteria. A source counted as relevant if it gave (i) a foundational technical specification or standard, (ii) peer-reviewed work on blockchain architecture or credential verification, (iii) research on OCR, document understanding or document forensics, or (iv) evidence about human oversight, access control, privacy or compliance workflows. We left a source out of the core synthesis if it had no clear link to these themes or lacked enough bibliographic detail to identify it.

Data extraction and synthesis. For every selected source, we noted its topic, what it contributes to the NOC-verification problem, and a limitation or implication for the proposed workflow. We then grouped findings by function: tamper-evident recording, off-chain storage, pre-approval document checks, forgery detection, officer accountability, and certificate expiry and renewal. The workflow we propose is a synthesis meant to guide future implementation. We are not claiming that an integrated system has been built or experimentally validated.

Limitations. A targeted narrative search cannot claim exhaustive coverage. We do not report a database-by-database search log, a count of duplicates removed, independent screening by several reviewers or a PRISMA flow diagram. Those details should be added only if the authors can show they were actually carried out.

I. Introduction

Hotels, commercial buildings, factories and other regulated businesses must hold No Objection Certificates (NOCs) and licenses before they open or keep operating. Government officers go through the submitted documents, approve or reject them, and keep the records in case of later inspection or investigation. In many places, this is still a manual, paper-based job. Verification means physical visits and repeated submissions. An approved certificate can be altered or forged afterwards. The approval history is hard for anyone outside the office to check. Expiry and renewal dates are chased by hand. This is as much a problem of trust as of efficiency: even a certificate that was approved correctly is difficult to prove genuine and unchanged years later [21].

Blockchain looks like a natural candidate for the job. A hash-chained ledger removes the need for one trusted record keeper [1], and programmable platforms add smart contracts that can hold approval logic [2]. Surveys cover the architecture, consensus mechanisms and uses of these systems [3]–[6], and a growing body of work looks at government services [18]–[20] and credential and certificate verification [22]–[29]. One pattern keeps coming up: store the full document off-chain and record only a cryptographic fingerprint on-chain [12], [13]. The fingerprint is computed with a collision-resistant function such as SHA-256 [14], not the broken SHA-1 [15], [16].

But a blockchain only preserves what was put into it. It cannot show that the entry was truthful in the first place [21]. Even choosing a blockchain at all needs a reason [5]. A second line of research applies AI to documents. OCR and document-understanding models pull out fields and structure [30]–[38], forgery-detection methods look for signs of manipulation [39]–[43], and human-factors studies warn that automated help can lead people to rely on it too much if it is poorly designed [44], [45]. The confidence scores shown to reviewers also have to be trustworthy [47].

Because these strands are converging, a synthesis focused on method is more useful here than a standard technology survey. We are less interested in which algorithms are popular and more interested in how different methods can be joined into one verification lifecycle, where AI assists, the officer decides, and the blockchain keeps the evidence. The paper makes four contributions: (i) a synthesis of 50 sources across blockchain, document AI, forgery detection and governance, organised by role; (ii) the identification of methodological fragmentation as the central gap; (iii) a set of requirements drawn from findings across the studies; and (iv) a modular path for choosing methods in future NOC verification systems.

II. Literature Review

A. Blockchain and Smart-Contract Foundations

Studies [1]–[11] supply the technical base. Nakamoto [1] introduced a hash-chained, proof-of-work ledger, and Buterin [2] extended the idea with smart contracts. Surveys and overviews [3], [4], [6] describe the architecture and where it is used. Wüst and Gervais [5] offer a decision scheme for judging when a blockchain is really justified and whether it should be public or permissioned. Hyperledger Fabric [7] is a widely cited permissioned platform that suits consortium settings such as networks linking several departments. Work on security and performance [8], attack taxonomies [9], program-analysis tools [10] and architectural taxonomies [11] shows that bugs in smart contracts and poor design choices can weaken the very records a system is supposed to protect. The common lesson is that a blockchain is only as strong as the contract code and architecture around it.

B. Off-Chain Storage, Hashing and Privacy

Studies [12]–[17] ask where data should live and how it should be fingerprinted. Content-addressed storage such as IPFS [12] and off-chaining patterns [13] support keeping large or sensitive files outside the ledger. The SHA-2 family is standardised in FIPS 180-4 [14], and practical collision attacks on SHA-1 [15], [16] explain why newer systems use SHA-256. Finck [17] looks at the clash between an immutable ledger and data-protection law, which implies that personal data should not be written to a public chain. The weakness of this group is that storage and hashing are treated as general infrastructure, not as parts of an approval workflow.

C. Blockchain in Government and Credential Verification

Studies [18]–[29] cover applications. Government-focused work [18]–[20] sees benefits for information sharing and transparency, but stresses governance, legal readiness and barriers to adoption. Lemieux [21] argues that a blockchain protects records after they are entered but cannot guarantee that a record was genuine when it was created. Education is the most developed area, with proposals and prototypes for recording achievements and credentials [22]–[27] and identity and credential standards [28], [29]. Together they show that hash-anchored verification is feasible. Most of these systems, however, verify issuance by an institution that is already known. Few deal with documents that come from third parties and need to be checked before approval.

D. AI-Based Document Extraction and Understanding

Studies [30]–[38] trace the move from classical OCR to learned document understanding. Tesseract [30] is still a widely used open-source baseline. Neural recognition and detection models [31]–[33] read text better, and layout-aware pretrained models [34]–[36] extract structured fields from document images, sometimes without a separate OCR step. Public benchmarks such as CORD [37] and FUNSD [38] cover receipts and forms, not government certificates. One finding is consistent across these works: extraction quality depends on scan quality and document type. Extracted fields should therefore be treated as evidence to check, not as ground truth.

E. Document Forgery Detection

Studies [39]–[43] deal with detecting manipulation. Classical image-forensics methods [39], [40] find copied regions and statistical inconsistencies. Document-specific methods [41] analyse texture patterns in printed documents, and benchmark efforts [42] test fraud detection on scanned documents. Verdoliva [43] notes that generalising to manipulations the detector has not seen is still hard. These methods are usually tested as stand-alone detectors, and they are rarely connected to an approval workflow or a tamper-evident record.

F. Human Oversight, Access Control and Compliance Lifecycle

Studies [44]–[50] look at how automation fits into institutional processes. Parasuraman and Manzey [44] document complacency and automation bias. Amershi et al. [45] give guidelines for human-AI interaction, including making system behaviour understandable and correctable. Wirtz et al. [46] review applications of AI in the public sector and the challenges they bring. Guo et al. [47] show that modern neural networks can be badly calibrated, so any confidence value shown to a user needs calibration first. Role-based access control [48] and signed-token standards such as JWT [49] provide the authorisation and authentication layer, and the RegTech literature [50] supports automated compliance monitoring. Expiry tracking and certificate renewal, by contrast, are rarely discussed in the blockchain credential literature.

III. Review Methodology

The review follows an analytical protocol built around methods. The candidate pool came from scholarly database queries (IEEE Xplore, ACM Digital Library, Elsevier ScienceDirect, Springer and Google Scholar), supplemented with standards and foundational papers wherever specific technical evidence was needed. The final analytical set has 50 sources covering blockchain foundations, off-chain storage and hashing, government and credential applications, document AI, forgery detection, and human oversight and compliance.

A. Search Strategy

Representative search combinations were:

- "blockchain" AND "certificate verification"
- "blockchain" AND "e-governance" AND "document"
- "smart contract" AND "document hash"
- "off-chain storage" AND blockchain
- "OCR" AND "document information extraction"
- "document forgery detection"
- "human-in-the-loop" AND "public sector" AND AI
- "role-based access control" AND blockchain
- "compliance" AND "expiry" AND "renewal"

B. Inclusion and Exclusion Criteria

We included a source if it made an empirical, methodological or standards contribution to blockchain record integrity, smart-contract design, off-chain storage, document extraction, forgery detection, human oversight of automated systems, access control or compliance management. We excluded vendor marketing, non-refereed commentary, and generic blockchain or AI overviews that had no bearing on document verification or public-sector workflows. Foundational and standards documents stayed in when they directly backed a design decision, such as the choice of hash function.

C. Data Extraction

Each selected source was catalogued by application context, functional role, core technique, main contribution, limitation and methodological implication. The aim was to turn limitations that kept recurring across different publications into a practical set of architectural requirements. The resulting 50-source matrix (Table II) is the basis for the research-gap and method-selection analysis.

D. Analytical Procedure

The synthesis ran in four phases: (1) sort the sources by their role in a verification lifecycle; (2) pull out the cross-cutting capabilities and failure points; (3) derive concrete methodological requirements that address those failure points; and (4) map candidate technologies against the requirements. Working in this order keeps the recommended architecture grounded in the literature instead of being fitted to it afterwards.

IV. Results and Discussion

A. Evolution of the Research Area

The 50 sources show a progression, not a pile of unrelated technologies:

1. Paper-based certificate review and manual record keeping
2. Digital portals and role-based access control
3. Hash-anchored, tamper-evident blockchain records
4. Smart-contract-based approval workflows
5. OCR and layout-aware document understanding
6. Forgery detection and calibrated confidence estimation
7. Human-in-the-loop, lifecycle-aware verification with expiry and renewal tracking

The field has moved from keeping records, to making them tamper-evident, to checking what is recorded before it is recorded, and finally to keeping the record valid over time.

B. Cross-Study Findings

Table I. Findings across the studies and the methodological requirements they point to

Hash anchoring gives tamper-evident records [1], [12], [24]	A blockchain cannot confirm a document was genuine when entered [21]	Check content before recording	AI-assisted preliminary validation
Off-chain storage with on-chain fingerprints is standard [12], [13]	Personal data on-chain conflicts with data-protection law [17]	Minimal on-chain data	Hash plus metadata on-chain, documents off-chain
SHA-256 is standardized; SHA-1 is broken [14]–[16]	Applications often leave the hash choice unstated	Collision-resistant fingerprint	SHA-256 document hash
Smart contracts encode workflow logic [2], [7]	Contract flaws can corrupt records [8]–[10]	Secure, auditable approval logic	Reviewed smart contract; permissioned option
OCR and document AI extract fields [30]–[38]	Quality varies with scan and document type	Treat output as evidence, not ground truth	OCR with consistency checks and flags
Forgery detectors exist [39]–[43]	Evaluated as isolated detectors	Link detection to approval	Suspicion flagging that feeds officer review
Automation can cause over-reliance [44], [45]	Risk of rubber-stamping AI output	Human accountability	Officer-authoritative decision
Neural confidence can mislead [47]	Uncalibrated scores mislead reviewers	Trustworthy confidence	Calibrated confidence indicators
Access control and signed tokens are mature [48], [49]	Rarely tied to approval roles on-chain	Attributable actions	RBAC, JWT and audit logging
Credential systems verify issuance [22]–[29]	Expiry and renewal rarely handled	Lifecycle management	Expiry alerts, renewal versioning, new hashes

Table II. Summary of the 50 reviewed sources

1	Nakamoto (2008)	Hash-chained, proof-of-work ledger with no trusted third party	Built for currency, not document governance; establishes the tamper-evidence principle
2	Buterin (2014)	Ethereum: Turing-complete smart contracts on a shared ledger	Cost and visibility on a public chain need assessment; enables on-chain approval logic
3	Zheng et al. (2017)	Overview of blockchain architecture, consensus and trends	General purpose; domain design is left to implementers
4	Casino et al. (2019)	Systematic review classifying blockchain applications by sector	Reports open issues such as scalability and privacy
5	Wüst and Gervais (2018)	Decision scheme for when a blockchain is justified, public vs. permissioned	Helps justify the design choice
6	Yaga et al. (2018)	NIST overview of blockchain properties and components	Descriptive, not prescriptive, for document systems
7	Androulaki et al. (2018)	Hyperledger Fabric, a permissioned execute-order-validate platform	Platform-specific; relevant to inter-departmental networks
8	Rouhani and Deters (2019)	Survey of smart-contract security, performance and applications	Contracts need a security review before deployment
9	Atzei et al. (2017)	Taxonomy of attacks on Ethereum smart contracts	Contract flaws can undermine recorded approvals
10	Luu et al. (2016)	Shows many contracts are exploitable; offers a symbolic-execution tool	Tool-specific; supports pre-deployment analysis
11	Xu et al. (2017)	Taxonomy of architectural decisions for blockchain-based systems	Generic; guides on-chain vs. off-chain decisions
12	Benet (2014)	IPFS: content-addressed peer-to-peer file system	Persistence and access control need extra mechanisms
13	Eberhardt and Tai (2017)	Patterns and trade-offs for off-chain data and computation	Supports keeping full documents off-chain, anchored by a hash

14	NIST FIPS 180-4 (2015)	Standard specifying the SHA-2 family, including SHA-256	A standard, not an application study; basis for fingerprinting
15	Wang et al. (2005)	Collision attack on full SHA-1	A reason to avoid SHA-1 for document fingerprints
16	Stevens et al. (2017)	First practical SHA-1 collision	Reinforces the preference for SHA-256
17	Finck (2019)	Analysis of the tension between immutability and GDPR	Keep personal data off-chain
18	Ølnes et al. (2017)	Benefits and implications of blockchain for government information sharing	Highlights governance and adoption challenges
19	Alketbi et al. (2020)	Reference model for blockchain government services (Dubai case)	Largely conceptual; limited evaluation
20	Kassen (2022)	Blockchain for automating public information processes in e-government	Institutional and legal readiness limit adoption
21	Lemieux (2016)	Examines whether blockchain makes records trustworthy	Cannot ensure a record was genuine at entry; motivates pre-approval validation
22	Sharples and Domingue (2016)	Blockchain-based permanent record of educational achievements	Proposal-level; no validation of document content
23	Grech and Camilleri (2017)	European Commission JRC report on blockchain in education	Notes privacy and standardization challenges
24	Turkanović et al. (2018)	EduCTX: blockchain-based higher-education credit platform	Credit-centric; not a certificate lifecycle
25	Gräther et al. (2018)	Blockchain-anchored lifelong learning passport prototype	Prototype; focus on learner-held credentials
26	Chen et al. (2018)	Review of blockchain's potential applications in education	Few evaluated deployments
27	Jirgensons and Kapenieks (2018)	Blockchain for credential assessment and management	Conceptual
28	Mühle et al. (2018)	Survey of self-sovereign identity components	Identity-focused; informs verifiable credentials
29	Sporny et al. (2019)	W3C Verifiable Credentials data model	A standard; adoption for regulatory certificates is a future step
30	Smith (2007)	Architecture of the open-source Tesseract OCR engine	Sensitive to image quality; extraction alone does not validate
31	Shi et al. (2017)	CRNN: end-to-end image-based sequence recognition	Scene-text focus; no consistency checking
32	Baek et al. (2019)	CRAFT: character-region-awareness text detection	Detection only
33	Li et al. (2023)	TrOCR: transformer-based OCR with pre-trained models	Compute-intensive
34	Xu et al. (2020)	LayoutLM: joint text-and-layout pre-training	Needs labeled data for each document type
35	Huang et al. (2022)	LayoutLMv3: unified text and image pre-training	Same data requirement; not tested on certificates
36	Kim et al. (2022)	Donut: OCR-free document understanding	Output reliability needs independent checking
37	Park et al. (2019)	CORD: receipt dataset for post-OCR parsing	Receipts, not government certificates

38	Jaume et al. (2019)	FUNSD: noisy scanned form understanding dataset	Small; forms rather than NOCs
39	Farid (2009)	Survey of image-forgery detection techniques	Mainly photographic images
40	Mahdian and Saic (2007)	Copy-move forgery detection with blur moment invariants	Targets copied regions only
41	Cruz et al. (2017)	Local binary patterns for document forgery detection	Limited to specific forgery types and datasets
42	Artaud et al. (2018)	Contest report on fraud detection in scanned documents	Benchmark-scale evaluation
43	Verdoliva (2020)	Overview of media forensics and deepfakes	Generalization to unseen manipulations is limited
44	Parasuraman and Manzey (2010)	Review of complacency and automation bias	AI flags must not replace officer judgment
45	Amershi et al. (2019)	Guidelines for human-AI interaction	General; informs how to explain and correct flags
46	Wirtz et al. (2019)	AI applications and challenges in the public sector	Conceptual; highlights governance and ethics
47	Guo et al. (2017)	Modern neural networks are miscalibrated; calibration remedies	Confidence scores need calibration before display
48	Sandhu et al. (1996)	Family of role-based access control models	Policy design remains application-specific
49	Jones et al. (2015)	JSON Web Token standard (RFC 7519)	Needs expiry and key management
50	Amer et al. (2017)	RegTech and the reconceptualization of financial regulation	Financial-sector focus; informs automated compliance monitoring

C. Research Gap

The reviewed sources show that each building block is well documented on its own: hash anchoring and smart contracts [1], [2], [12], [13], credential verification prototypes [22]–[29], OCR and document understanding [30]–[38], forgery detection [39]–[43], access control [48], [49] and compliance monitoring [50]. The open problem is how they fit together, which we call methodological fragmentation. In the reviewed corpus, credential-verification systems anchor documents that are simply assumed to be genuine. Document-AI work stops at extraction or detection. Research on human oversight is seldom tied to either. We found few architectures that do all of the following at once: (a) check an incoming certificate before approval, (b) keep the authorised officer as the final decision-maker, (c) record a fingerprint and minimal approval metadata on a tamper-evident ledger, (d) allow later independent re-verification, and (e) manage expiry, renewal and versioning across the certificate's life. Long-term, end-to-end evaluation of such a workflow on regulatory certificates is also scarce.

The research gap, then, is this. Blockchain hash anchoring, document AI, forgery detection and access control have each proved useful alone, but their integration into a unified, officer-authoritative, lifecycle-aware NOC verification methodology is still fragmented and insufficiently evaluated.

D. Methodological Requirements Derived from the Gap

Table III. Methodological requirements derived from the literature

Check documents before approval	OCR extraction and consistency validation (certificate number, entity name, issuer, dates)	[30]–[38], [21]
Detect suspicious submissions	Suspicion flagging for unreadable, incomplete or altered fields	[39]–[43]
Trustworthy AI assistance	Calibrated confidence indicators; flags that are explained and correctable	[45], [47]
Human accountability	Officer-authoritative approve, reject or correct decision; AI never grants approval	[44], [45]

Tamper-evident fingerprint	SHA-256 hash of the approved version	[14]–[16]
Immutable approval record	Smart contract storing hash, record ID, timestamp and decision metadata	[2], [7]–[10]
Privacy and storage separation	Full documents and personal data off-chain; minimal data on-chain	[12], [13], [17]
Independent verification	Re-hash the presented document and compare with the on-chain record	[18], [24]
Attributable actions	Role-based access control, JWT authentication, audit logging	[48], [49]
Lifecycle management	Expiry alerts, renewal versioning with a new hash per version, compliance analytics	[50]
Platform justification	Decision between a public and a permissioned ledger	[5], [7]

E. Path Toward Method Selection

The literature suggests that closing this gap calls for a modular, multi-tier architecture, not a single monolithic algorithm. Pre-approval validation needs an OCR and consistency-checking tier whose output is handed to the officer as evidence. Tamper evidence needs SHA-256 fingerprinting and a smart contract that records the approved version and its metadata. Privacy and scalability need off-chain document storage, with only minimal data on-chain. Accountability needs role-based access control and audit logging. Long-term validity needs a compliance engine that tracks expiry and sends renewal reminders, with each renewal producing a new document version and a new hash.

The resulting chain of methods is:

Entity upload → OCR extraction → consistency validation and flags → officer decision → SHA-256 hashing → on-chain record → later re-verification and audit → expiry tracking → renewal and new version

F. Proposed Methodological Direction for Future Research

Based on this synthesis, future work should favour an integrated pipeline over isolated verification or isolated AI components. The submission tier should register documents through a role-based portal. The validation tier should extract key fields and run consistency checks against the submission and existing records, returning a structured result with calibrated confidence indicators. The decision tier should keep the authorised officer as the only approving authority. The evidence tier should hash the approved version with SHA-256 and write the hash, together with minimal approval metadata, to a smart contract. The verification tier should let an auditor re-hash a presented document and compare it with the on-chain record: a match indicates an unmodified document, and a mismatch triggers an investigation. The lifecycle tier should send renewal reminders ahead of expiry, mark expired certificates, and record each renewal as a new version.

Two design choices need particular care. First, writing an officer's identity to a public ledger raises the data-protection concerns discussed in [17], so a permissioned network [7] or pseudonymous identifiers should be evaluated, guided by the decision framework in [5]. Second, uncalibrated confidence and automation bias can mislead reviewers [44], [47], so the interface should show each flag with its reasons and let the officer correct or override it [45].

Future implementations should be measured on: field-level OCR accuracy for the certificate types in scope; precision and recall of the suspicion flags on realistic and synthetic tampering; officer review time with and without AI assistance; tamper-detection rate at re-verification; transaction cost and latency; and how timely renewals are. Extensions worth studying include integration with government portals, a multi-department permissioned network, machine-learning-based fraud detection, and signed portable credentials based on the W3C Verifiable Credentials model [29].

V. Conclusion

This review brought together 50 sources on blockchain foundations, off-chain storage and hashing, government and credential applications, document AI, forgery detection, and human oversight and compliance. Individual capabilities are clearly mature. Hash-anchored records are well understood, OCR and document-understanding models can extract structured fields, forgery detectors exist, and access-control and token standards are established.

The main obstacle we found is methodological fragmentation, not a shortage of algorithms. A blockchain cannot confirm that an entered record was genuine. Document AI is rarely tied to an approval decision. Certificate expiry and renewal are seldom addressed at all. To move the field forward, next-generation NOC verification architectures should combine AI-assisted preliminary validation, officer-authoritative approval, SHA-256 fingerprints with on-chain approval metadata, off-chain document storage, role-based access control, and automated expiry and renewal versioning in a single closed loop. Future work needs to test empirically whether this integration improves validation accuracy, tamper-detection reliability, officer efficiency and renewal compliance.

REFERENCES

- [1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008.
- [2] V. Buterin, "A next-generation smart contract and decentralised application platform" (Ethereum white paper), 2014.
- [3] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, "An overview of blockchain technology: Architecture, consensus, and future trends," in Proc. IEEE Int. Congress on Big Data, 2017.
- [4] F. Casino, T. K. Dasaklis, and C. Patsakis, "A systematic literature review of blockchain-based applications: Current status, classification and open issues," *Telematics and Informatics*, vol. 36, 2019.
- [5] K. Wüst and A. Gervais, "Do you need a blockchain?" in Proc. Crypto Valley Conf. on Blockchain Technology, 2018.
- [6] D. Yaga, P. Mell, N. Roby, and K. Scarfone, "Blockchain technology overview," NISTIR 8202, 2018.
- [7] E. Androulaki et al., "Hyperledger Fabric: A distributed operating system for permissioned blockchains," in Proc. EuroSys, 2018.
- [8] M. Rouhani and R. Deters, "Security, performance, and applications of smart contracts: A systematic survey," *IEEE Access*, vol. 7, 2019.
- [9] N. Atzei, M. Bartoletti, and T. Cimoli, "A survey of attacks on Ethereum smart contracts (SoK)," in Proc. Principles of Security and Trust (POST), 2017.
- [10] L. Luu, D.-H. Chu, H. Olickel, P. Saxena, and A. Hobor, "Making smart contracts smarter," in Proc. ACM CCS, 2016.
- [11] X. Xu et al., "A taxonomy of blockchain-based systems for architecture design," in Proc. IEEE Int. Conf. on Software Architecture (ICSA), 2017.
- [12] J. Benet, "IPFS – Content-addressed, versioned, P2P file system," arXiv:1407.3561, 2014.
- [13] S. Eberhardt and S. Tai, "On or off the blockchain? Insights on off-chaining computation and data," in Proc. ESOC, 2017.
- [14] NIST, "Secure Hash Standard (SHS)," FIPS PUB 180-4, 2015.
- [15] X. Wang, Y. L. Yin, and H. Yu, "Finding collisions in the full SHA-1," in Proc. CRYPTO, 2005.
- [16] M. Stevens, E. Bursztein, P. Karpman, A. Albertini, and Y. Markov, "The first collision for full SHA-1," in Proc. CRYPTO, 2017.
- [17] M. Finck, "Blockchain and the General Data Protection Regulation: Can distributed ledgers be squared with European data protection law?" European Parliamentary Research Service, 2019.
- [18] S. Ølne, J. Ubacht, and M. Janssen, "Blockchain in government: Benefits and implications of distributed ledger technology for information sharing," *Government Information Quarterly*, vol. 34, no. 3, 2017.
- [19] A. Alketbi, Q. Nasir, and M. A. Talib, "Novel blockchain reference model for government services: Dubai government case study," *Int. J. Syst. Assur. Eng. Manag.*, 2020.
- [20] M. Kassen, "Blockchain and e-government innovation: Automation of public information processes," *Information Systems*, vol. 103, 2022.
- [21] V. L. Lemieux, "Trusting records: Is blockchain technology the answer?" *Records Management Journal*, vol. 26, no. 2, 2016.
- [22] M. Sharples and J. Domingue, "The blockchain and kudos: A distributed system for educational record, reputation and reward," in Proc. EC-TEL, 2016.
- [23] A. Grech and A. F. Camilleri, "Blockchain in education," European Commission Joint Research Centre, 2017.
- [24] M. Turkanović, M. Höbl, K. Košič, M. Heričko, and A. Kamišalić, "EduCTX: A blockchain-based higher education credit platform," *IEEE Access*, vol. 6, 2018.
- [25] P. Gräther et al., "Blockchain for education: Lifelong learning passport," in Proc. ERCIM Blockchain Workshop, 2018.
- [26] G. Chen, B. Xu, M. Lu, and N.-S. Chen, "Exploring blockchain technology and its potential applications for education," *Smart Learning Environments*, vol. 5, 2018.
- [27] A. Jirgensons and J. Kapenieks, "Blockchain and the future of digital learning credential assessment and management," *Journal of Teacher Education for Sustainability*, vol. 20, no. 1, 2018.
- [28] A. Mühle, A. Grüner, T. Gayvoronskaya, and C. Meinel, "A survey on essential components of a self-sovereign identity," *Computer Science Review*, vol. 30, 2018.
- [29] M. Sporny, D. Longley, and D. Chadwick, "Verifiable Credentials Data Model 1.0," W3C Recommendation, 2019.
- [30] R. Smith, "An overview of the Tesseract OCR engine," in Proc. ICDAR, 2007.
- [31] B. Shi, X. Bai, and C. Yao, "An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 11, 2017.

- [32] Y. Baek, B. Lee, D. Han, S. Yun, and H. Lee, "Character region awareness for text detection," in Proc. IEEE/CVF CVPR, 2019.
- [33] M. Li et al., "TrOCR: Transformer-based optical character recognition with pre-trained models," in Proc. AAAI, 2023.
- [34] Y. Xu et al., "LayoutLM: Pre-training of text and layout for document image understanding," in Proc. ACM KDD, 2020.
- [35] Y. Huang, T. Lv, L. Cui, Y. Lu, and F. Wei, "LayoutLMv3: Pre-training for document AI with unified text and image masking," in Proc. ACM Multimedia, 2022.
- [36] G. Kim et al., "OCR-free document understanding transformer," in Proc. ECCV, 2022.
- [37] S. Park et al., "CORD: A consolidated receipt dataset for post-OCR parsing," Workshop on Document Intelligence at NeurIPS, 2019.
- [38] G. Jaume, H. K. Ekenel, and J.-P. Thiran, "FUNSD: A dataset for form understanding in noisy scanned documents," in Proc. ICDAR-OST, 2019.
- [39] H. Farid, "Image forgery detection," IEEE Signal Processing Magazine, vol. 26, no. 2, 2009.
- [40] B. Mahdian and S. Saic, "Detection of copy-move forgery using a method based on blur moment invariants," Forensic Science International, vol. 171, no. 2-3, 2007.
- [41] V. Cruz, J. Sidère, M. Coustaty, and J.-M. Ogier, "Local binary patterns for document forgery detection," in Proc. ICDAR, 2017.
- [42] C. Artaud, N. Sidère, A. Doucet, J.-M. Ogier, and V. Poulain D'Andecy, "Find it! Fraud detection contest report," in Proc. ICPR, 2018.
- [43] L. Verdoliva, "Media forensics and DeepFakes: An overview," IEEE J. Sel. Topics Signal Process., vol. 14, no. 5, 2020.
- [44] R. Parasuraman and D. H. Manzey, "Complacency and bias in human use of automation: An attentional integration," Human Factors, vol. 52, no. 3, 2010.
- [45] S. Amershi et al., "Guidelines for human-AI interaction," in Proc. ACM CHI, 2019.
- [46] B. W. Wirtz, J. C. Weyerer, and C. Geyer, "Artificial intelligence and the public sector—Applications and challenges," Int. J. Public Administration, vol. 42, no. 7, 2019.
- [47] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in Proc. ICML, 2017.
- [48] R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman, "Role-based access control models," IEEE Computer, vol. 29, no. 2, 1996.
- [49] M. Jones, J. Bradley, and N. Sakimura, "JSON Web Token (JWT)," RFC 7519, IETF, 2015.
- [50] D. W. Arner, J. Barberis, and R. P. Buckley, "FinTech, RegTech, and the reconceptualisation of financial regulation," Northwestern J. Int. Law & Business, vol. 37, no. 3, 2017.