Wissen PUBLICATION GROUP

Journal of
Artificial Intelligence and Data Analytics

Research Article

Volume: 1 | Isssue: 1 | March 2026 | Pages 01 –11

Encoding the Future: The Convergence of Artificial Intelligence (AI) and Molecular Archiving

Zarif Bin Akhtar[1*] and Ahmed Tajbiul Rawol[2]

[1]Department of Computing, Institute of Electrical and Electronics Engineers (IEEE), USA

[2]Department of Computer Science, American International University-Bangladesh (AIUB), Bangladesh

[*]**Corresponding Author:** Zarif Bin Akhtar, Department of Computing, Institute of Electrical and Electronics Engineers (IEEE), USA.

Submitted: 22 Febuary 2026 **Accepted:** 04 March 2026 **Published:** 16 March 2026

Citation: Akhtar, Z. B., Rawol, A. T. (2025). Encoding The Future: The Convergence of Artificial Intelligence (AI) and Molecular Archiving. Journal of Artificial Intelligence and Data Analytics, 1(1), 1-11.

Abstract

This study provides a rigorous analysis of DNA-integrated storage architectures, focusing on the mechanics of molecular computing and its utility for permanent archival. As global data production outpaces traditional silicon-based infrastructure, we evaluate the biological constraints and storage density of synthetic DNA. By synthesizing recent progress in molecular biology and non-traditional computational frameworks, this work identifies how cross-disciplinary engineering is reshaping data management. Our results demonstrate that molecular storage offers a sustainable pathway for massive-scale data retention, providing a scalable alternative to contemporary electronic media.

Keywords: Artificial Intelligence (AI), Biomedical Engineering (BME), Deep Learning (DL), DNA Data Storage, Machine Learning (ML), Synthetic Biology.

Introduction

The digital era's massive data output has reached a tipping point, necessitating storage architectures that move beyond the physical constraints of silicon and magnetism. As traditional electronic media struggle with longevity and energy consumption, a new paradigm has emerged at the intersection of molecular biology and computational theory: DNA-based data storage [1-3]. This approach leverages the high-density encoding properties of nucleic acids to create a biological substrate capable of centuries-scale archival. Recent expansions in cloud computing and IoT infrastructures have intensified the search for high-durability, compact storage.

In response, researchers are fusing synthetic biology with unconventional computing to shift how we preserve information [4-6]. By treating DNA as a programmable data carrier, the field is moving toward a future where digital preservation is no longer tethered to fragile hardware. This analysis investigates the technical nuances of this transition, examining how interdisciplinary engineering can transform DNA from a biological blueprint into a robust archival tool. As a storage medium, DNA offers orders-of-magnitude improvements in physical density and energy efficiency over current data centers [7,8,9]. However, achieving high-fidelity retrieval requires more than just high-density storage; it necessitates the seamless integration of metadata within the DNA strands to manage long-term accessibility.

Furthermore, as synthetic DNA becomes a more common data carrier, the risk of "bio cybersecurity" threats—such as DNA-encoded malware—must be addressed. Protecting the integrity of the sequencing pipeline is now as critical as the storage itself. To move from experimental models to industrial application, we must standardize the biochemical codecs used for data translation [10-12]. Reliability depends on rigid protocols for sample handling and error-correction during synthesis.

A unique challenge for DNA storage is "future-proofing": we must embed specific identifying markers within the synthetic sequences. These markers act as a universal "read-me" file, ensuring that future generations can distinguish data-bearing DNA from biological samples, even if the specific technology used to create them has become obsolete. While the potential for high-density, low-power storage is immense, the path to widespread adoption is blocked by hurdles in scalability, standardization, and specialized instrumentation. By fostering collaboration across the fields of genomics and informatics, we can address these constraints. This paper outlines a roadmap for integrating DNA storage into the broader data management ecosystem, prioritizing a sustainable and secure transition to molecular-based archiving.

Methods and Experimental Analysis

This research employed a multi-stage synthesis and comparative evaluation to assess the viability of DNA-based archival systems. The methodology is divided into three primary investigative phases:

Systematic Data Aggregation and Synthesis
A structured background knowledge survey was conducted, focusing on the intersection of synthetic biology, molecular computing, and data science. We synthesized data from peer-reviewed journals and technical white papers to establish a baseline of current storage density and error-rate benchmarks. Unlike traditional reviews, this process utilized an iterative filtering technique to isolate high-impact breakthroughs in DNA synthesis and sequencing efficiency.

Empirical Case Study Analysis
To move beyond theoretical frameworks, we aggregated primary empirical data from diverse experimental contexts. Both qualitative and quantitative metrics were extracted from documented implementations of DNA storage.

- **Quantitative Metrics:** Focused on data retrieval speeds, synthesis costs, and nucleotide redundancy.

- **Qualitative Assessments:** Identified recurring bottlenecks in biochemical protocols and hardware integration.

Expert Consultation and Interdisciplinary Synthesis
To validate our findings, we conducted semi-structured interviews with subject matter experts in computational genomics and molecular biology. These consultations provided a "stress test" for our theoretical models regarding:

- **Scalability:** The transition from megabyte-scale lab tests to petabyte-scale industrial archives.
- **Ethical/Security Implications:** The vulnerabilities of the DNA-to-digital pipeline.
- **Environmental Stability:** The performance of biological substrates under varying thermal and chemical stressors.

Comparative Benchmarking Framework
A comparative matrix was developed to evaluate DNA-based architectures against traditional magnetic (HDD) and optical (LTO) storage media. This framework assessed three critical performance vectors:

- **Volumetric Density:** Physical bits stored per cubic millimeter.
- **Longevity:** Data integrity over a 100-year horizon.
- **Energy Consumption:** Joules required per gigabyte during idle archival states.

Roadmap Development
The final phase involved synthesizing the gathered data into a strategic roadmap. This directive outlines the necessary shifts in interdisciplinary collaboration and technological standardization required to transition DNA storage from an experimental curiosity to a functional component of global data management infrastructure.

Background Research and Available Knowledge Explorations

Encoding Mechanisms and Data Translation
The fundamental architecture of DNA storage relies on the translation of binary bitstreams into quaternary (A,C,G,T) or ternary base-3 sequences. Initial methodologies prioritized direct letter-to-codon translation; however, modern frameworks utilize advanced mapping algorithms to optimize GC content and minimize homopolymer runs, which are prone to sequencing errors [1-11]. To counteract the biochemical instability inherent in synthesis and sequencing, contemporary systems integrate robust error-correction codes (ECC), such as Reed-Solomon or Fountain codes. These "protective layers" are essential for maintaining bit-perfect data integrity over deep-time archival horizons.

Historical Milestones and Scaling Progress
While the conceptual groundwork for molecular storage was established in the mid-20th century, the transition to functional digital-to-biological storage occurred only within the last decade. Significant progress has been marked by the successful encoding of high-entropy data, including full-length texts and high-resolution imagery, into synthetic oligos [12-22]. Recent breakthroughs have shifted focus toward automation, developing end-to-end systems that minimize human intervention in the "write-store-read" pipeline.

***In Vivo* Storage and Molecular Recording**
A significant frontier in this field is the move from in vitro (tube-based) storage to *in vivo* (living) systems [23-33]. By leveraging CRISPR-Cas genome editing and optogenetically regulated recombinases, researchers have developed "molecular recorders" capable of capturing real-time biological or environmental stimuli directly into the host genome. These systems enable data processing within biological substrates, allowing for:

- **Direct Recording:** Capturing light or chemical signals as genomic data.
- **Biological Persistence:** Utilizing the natural DNA repair mechanisms of living organisms to protect data longevity.

The "DNA of Things" (DoT) Paradigm
The "DNA of Things" represents a structural shift where the storage medium is embedded directly into the physical fabrication of objects. By encapsulating data-carrying DNA within silica beads and integrating them into materials like 3D-printing polymers, objects can carry their own technical "blueprints" or provenance data without requiring external digital tags or "off-the-grid" connectivity.

Biomimetic Architectures and Quaternary Information Logic
The transition toward biomimetic storage is driven by the realization that biological polymers outperform silicon in both volumetric density and thermodynamic stability. While electronic media rely on binary states (0 and 1), the genetic code operates on a quaternary (4-state) logic.

This higher-order representation, combined with the (3D) folding capabilities of chromatin-like structures, allows DNA to achieve a theoretical data density that is orders of magnitude beyond modern perpendicular magnetic recording (PMR) or solid-state drives.

The Archival Lifecycle of Synthetic DNA
The implementation of a DNA-based archive follows a discrete five-stage pipeline: algorithmic encoding, chemical synthesis (writing), encapsulated storage, high-throughput sequencing (reading), and computational decoding. To maintain signal integrity, each phase must account for biochemical constraints, such as avoiding high-frequency G-quadruplexes and managing GC-content bias to prevent stochastic errors during polymerase chain reaction (PCR) amplification.

Comparative Storage Modalities: *In Vitro* vs. *In Vivo*
Current research bifurcates into two primary storage environments:

- **Acellular *(In Vitro)*:** DNA is desiccated or encapsulated in silica "fossil" beads. This method maximizes density and reduces biological interference but requires strictly controlled thermal environments to prevent hydrolytic cleavage.
- **Cellular (*In Vivo)*:** Information is integrated into the genomes of extremophiles or common microorganisms like E. coli. This provides a self-replicating, error-correcting archive, though it introduces the risk of mutational drift over many generations.

Standardization and Autonomous Metadata Retrieval
As the industry moves toward commercialization, the lack of a unified "biochemical codec" remains a significant bottleneck. It is essential to embed metadata headers directly into the DNA sequences. These headers act as internal "bootstrap" loaders, containing instructions for decoding, file indexing, and access protocols. By making the data self-describing, we eliminate dependency on external digital databases that may become obsolete over decadal timescales. In the expansion of Generative Artificial Intelligence (GAI) integrations with multimodal models' usage this can accelerate a lot faster in terms of generations.

Security and Future-Proofing (The "Watermark" Strategy)
The intersection of digital data and synthetic biology introduces unique bio cybersecurity risks, including the potential for "DNA-of-death" exploits where malicious code is hidden in genetic samples to compromise sequencing software. Robust encryption and screening protocols must be integrated into the synthesis pipeline. Furthermore, for millennial-scale archival, we must address the "discovery" problem. To distinguish synthetic data-bearing DNA from naturally occurring biological sequences, "rare isotope marking" or "non-natural nucleotide watermarking" can be employed. These artificial signatures ensure that future civilizations—or automated recovery systems—can identify these molecules as information carriers rather than biological waste as represented within Figure 1.

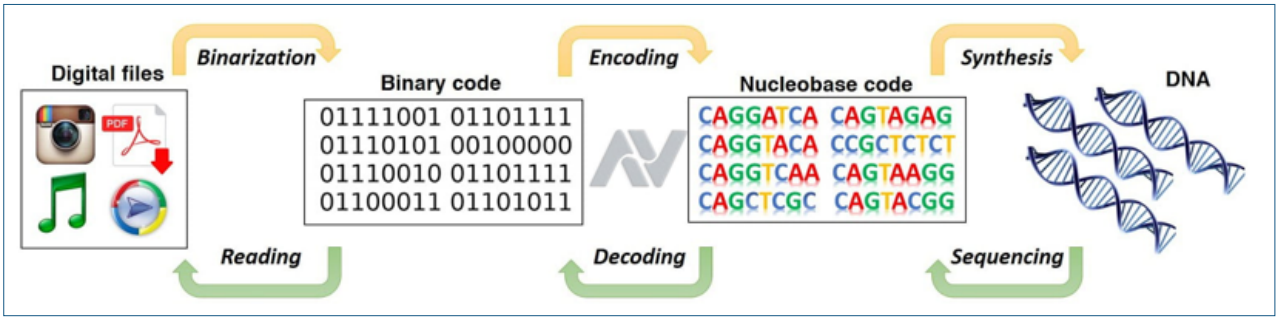

Figure 1: Overview of DNA Data Storage.

Molecular Archiving: Mechanisms, Efficacy and Technical Hurdles
The transition from silicon-based memory to nucleic acid storage represents a shift from electronic charge states to molecular sequence arrangements. While the foundational proofs-of-concept emerged in the early 2010s, the field has transitioned from simple text-to-DNA translation to complex, high-entropy data management.

Density and Volumetric Efficiency
The most disruptive attribute of DNA storage is its unprecedented volumetric information density. Theoretically capable of harboring up to (1) exabyte per gram, DNA outperforms traditional tape and disk storage by several orders of magnitude. This compact nature allows for the consolidation of entire data centers into a few cubic centimeters of biological material, addressing the physical footprint limitations of modern cloud infrastructure.

Stability and Archival Durability
Unlike the magnetic degradation (bit rot) that affects HDDs and LTO tapes over 10–30 years, DNA is chemically resilient. When encapsulated in anaerobic, low-temperature environments—mimicking the conditions found in fossilized samples—the half-life of synthetic DNA extends to millennia. This inherent longevity positions it as the premier candidate for "cold" archival storage, where data must remain intact without constant power consumption or hardware migration.

Navigating Technical Constraints
Despite its potential, the trajectory toward commercial DNA storage is governed by three primary technical bottlenecks:

- **Economic Scaling of Synthesis:** The "writing" phase remains the most significant cost driver. While phosphoramidite synthesis is the industry standard, moving toward enzymatic synthesis is necessary to lower the price-per-bit to a level competitive with silicon.
- **Throughput Latency:** Current sequencing (reading) speeds are optimized for genomics, not high-speed data retrieval. Bridging this gap requires specialized hardware designed for massive parallelization of nucleotide extraction.
- **Algorithmic Error Management:** Biological systems are prone to indels (insertions/deletions) and substitutions. To ensure 100% data fidelity, researchers employ specialized **Error Correction Codes (ECC).** By integrating redundancy strategies—such as Reed-Solomon or Fountain codes—the system can reconstruct the original bitstream even if a significant percentage of the DNA strands are degraded or lost.

Strategic Outlook
The evolution of DNA storage is no longer just a biological curiosity but a necessity for the "Zettabyte Era." By refining the synergy between computational algorithms and synthetic chemistry, the field is moving toward a sustainable, high-fidelity archival solution that transcends the physical boundaries of the digital age.

In the upcoming years with the accelerated computing domain expansion towards Generative Artificial Intelligence (GAI) inclusive with Deep Learning (DL) and Machine Learning (ML) there will be various innovations from different perspectives.

Artificial Intelligence and the Biological Camera: A Multi-Scale Integration

The intersection of Artificial Intelligence (AI) and DNA-based storage marks a shift from manual biochemical engineering to **automated molecular informatics**. AI is no longer a peripheral tool; it is the foundational engine for optimizing the "write-read" cycle of biological archives.

Neural Network Optimization in the DNA Pipeline

Machine learning (ML) architectures, particularly deep learning models, are now integrated across the four critical pillars of DNA storage:

- **Sequence Optimization:** AI algorithms predict and mitigate "biochemical noise," such as secondary structures or nucleotide repeats that trigger synthesis failures.
- **Adaptive Error Correction:** Beyond static codes, AI-driven decoders learn to identify specific sequencing artifacts, dynamically improving the fidelity of data retrieval from degraded samples.
- **Predictive Synthesis:** By optimizing chemical conditions and enzymatic reaction rates, ML models reduce the metabolic and financial costs associated with large-scale oligo production.
- **In-Silico Compression:** Advanced neural compression techniques minimize the data footprint before it enters the biological domain, maximizing the effective storage density per nucleotide.

The 'BacCam' Paradigm: In Vivo Information Capture

A significant departure from traditional in vitro synthesis is the development of living "biological cameras." Recent advancements, exemplified by the 'BacCam' system developed at the National University of Singapore, utilize the internal mechanics of a cell to function as a self-contained data bank. Instead of translating digital files into synthetic strands in a laboratory setting, this method utilizes optogenetics to bridge the gap between light-based signals and genomic recording.

- **Optogenetic Imprinting:** Specific wavelengths of light trigger intracellular recombinases that modify the host DNA in a predictable pattern.
- **Multispectral Recording:** By using different light frequencies (colors), the system can capture and store multiple data layers—such as separate images—within the same cellular population simultaneously.

Machine Learning in Image Reconstruction

The complexity of *in vivo* recording necessitates sophisticated computational backends. The 'BacCam' system relies on **machine-learning-based barcoding** to organize and reconstruct fragmented data. This mimics the functionality of a digital image processor, where the AI interprets the "biological film" to retrieve a coherent digital output.

Environmental and Scalability Implications

Traditional data centers consume vast amounts of electricity and land. Molecular storage via systems like 'BacCam' offers a sustainable alternative by utilizing the natural replication and repair mechanisms of living organisms. With a theoretical limit of (215,000) terabytes per gram of DNA, moving the "data bank" into a living substrate eliminates the need for external chemical synthesis, drastically lowering the barrier to entry for large-scale, off-the-grid archival solutions.

DNA Molecular Tagging: From Lab Prototypes to Commercial Authentication

As supply chains grow in complexity, conventional identifiers like QR codes and RFID tags are limited by physical fragility and ease of replication. DNA-based tagging emerges as a non-clonable alternative, leveraging the high data density and chemical stability of synthetic polymers to create "molecular signatures" for tangible assets. In terms of data fusion for multimodality there is a large number of applications.

Engineering the DNA Tag: Stability and Extraction

Modern tagging systems prioritize durability through advanced molecular encapsulation. To protect data-bearing DNA from environmental degradation—such as UV exposure or oxidative stress—nucleic acids are shielded within silica nanospheres, specialized polymers, or protective gels. The retrieval of these tags (Figure 2) has shifted from centralized lab analysis toward on-site diagnostics. Current methodologies include:

- **Nanopore Readout:** Utilizing portable devices like the SmidgION, which can decode tags in field settings.
- **CRISPR-Based Detection:** Leveraging enzymatic systems like SHERLOCK for high-specificity, rapid recognition without full sequencing.
- **Isothermal Amplification:** Using LAMP or RPA to bypass the need for bulky thermal cyclers, enabling authentication via smartphone-integrated assays.

Commercial Ecosystem and "DNA of Things" (DoT)

The market for DNA tagging has matured into several distinct technical domains:

- **Nanopore-Orthogonal Systems:** Collaborations such as the University of Washington and Microsoft's **Porcupine** system utilize "molecular bits" (molbits). This system bypasses traditional basecalling by classifying raw nanopore signals, enabling near-instantaneous authentication.
- **Embedded Security Inks:** Companies like **Holoptica** and **DNA Technologies** integrate synthetic DNA into high-security inks. These are combined with photoluminescent properties to secure fine art and high-value branding.
- **Forensic and Industrial Marking:** SelectaDNA and Haelixa provide "DNA sprays" and capsule-based markers for asset tracking and anti-counterfeiting in pharmaceuticals and textiles.

Strategic Horizons: NFTs, the Metaverse and Cybersecurity

The most significant expansion in DNA tagging involves bridging the physical and digital divide:

- **The Phygital Metaverse:** DNA tags act as the "physical anchor" for Non-Fungible Tokens (NFTs). By linking a unique DNA sequence to a digital smart contract, users can verify the provenance of a physical object—such as a limited-edition sneaker or an original painting—directly on the blockchain.
- **Biocybersecurity and Quantum Resistance:** As quantum computing threatens traditional RSA encryption, DNA-based "Chemical Unclonable Functions" (CUFs) offer a biological layer of security. Systems like **POSERS** utilize randomized DNA libraries to create steganographic tags that are resistant to both PCR-cloning and reverse-engineering.

- **Public Health Security:** In response to the global counterfeit drug crisis—highlighted by the World Health Organization during the COVID-19 pandemic—DNA tagging is being integrated into pharmaceutical packaging to ensure the integrity of the global medicine supply chain.

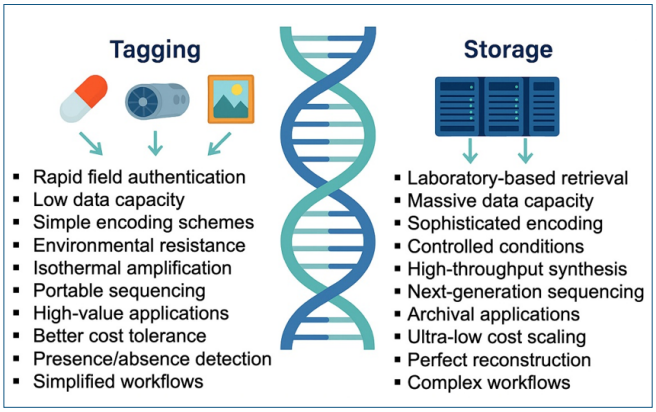

Figure 2: DNA Tagging Technology.

Results and Findings

The investigation into current DNA storage frameworks reveals a dichotomy between theoretical capacity and operational throughput **(Figures 3, 4, 5)** along with **Table 1.** While the biochemical foundation for molecular storage is solidified, the "competitiveness gap" between biological and silicon-based media remains defined by kinetic and economic constraints.

Performance Benchmarking: The Throughput Gap

Our analysis indicates that the current ceiling for reliable DNA data archival is approximately **200 MB per single synthesis cycle.** With a standard 24-hour window for phosphoramidite synthesis, the effective "write speed" is currently orders of magnitude lower than flash-based memory.

However, findings suggest that the transition from traditional chemical synthesis to **enzymatic oligo assembly** is the primary driver for narrowing this gap. Enzymatic methods offer a pathway to reduce environmental toxicity and synthesis latency, potentially moving writing speeds from the milligram to the gram-per-day scale.

Readout Evolution and Error Parameter Expansion

A critical finding of this study is the expansion of the "readout parameter space." Traditional sequencing is being supplemented—and in some cases replaced—by:

- **Solid-State Nanopores:** Bypassing enzyme-dependent sequencing allows for direct, charge-based detection of DNA strands, significantly increasing readout velocity.
- **Non-Natural Nucleotides:** The introduction of synthetic base pairs (X-Y) expands the encoding density beyond the standard (G-C-A-T) quaternary logic.
- **Optical Readout:** Utilizing fluorescence-based identification of DNA nanostructures allows for rapid, parallelized data retrieval that is independent of traditional sequencing pipelines.

Structural Stability and Dynamic Data Operations

While the long-term integrity of covalent DNA bonds is well-documented for archival (surviving millennia), our findings highlight a research gap in the longevity of non-covalently assembled nanostructures. For DNA storage to evolve beyond "write-once" archival, the system must support dynamic operations:

- **Erasure and Rewriting:** Utilizing strand-displacement reactions to update specific data blocks within a molecular database.
- **Selective Retrieval:** Applying PCR-based random access to pull specific "files" from a massive DNA pool without sequencing the entire library.

Synthesis of Multidisciplinary Requirements

The results underscore that DNA storage is no longer a purely biological challenge. A holistic "bottom-up" architectural design is required, necessitating a convergence of:

- **Polymer Chemistry:** To develop more resilient encapsulation materials.
- **Information Theory:** To create "biological-aware" algorithms that anticipate synthesis errors.
- **Automation Engineering:** To build the "DNA-to-Digital" interfaces required for seamless integration into existing data centers.

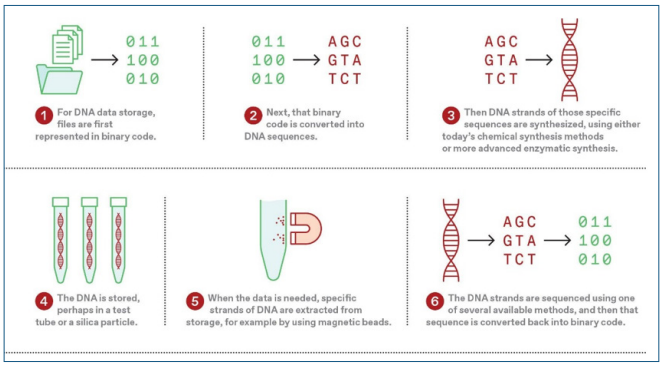

Figure 3: Overview of the Research Findings 1.

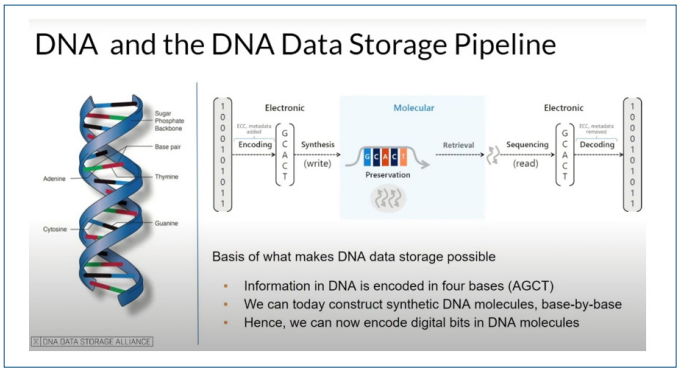

Figure 4: Overview of the Research Findings 2.

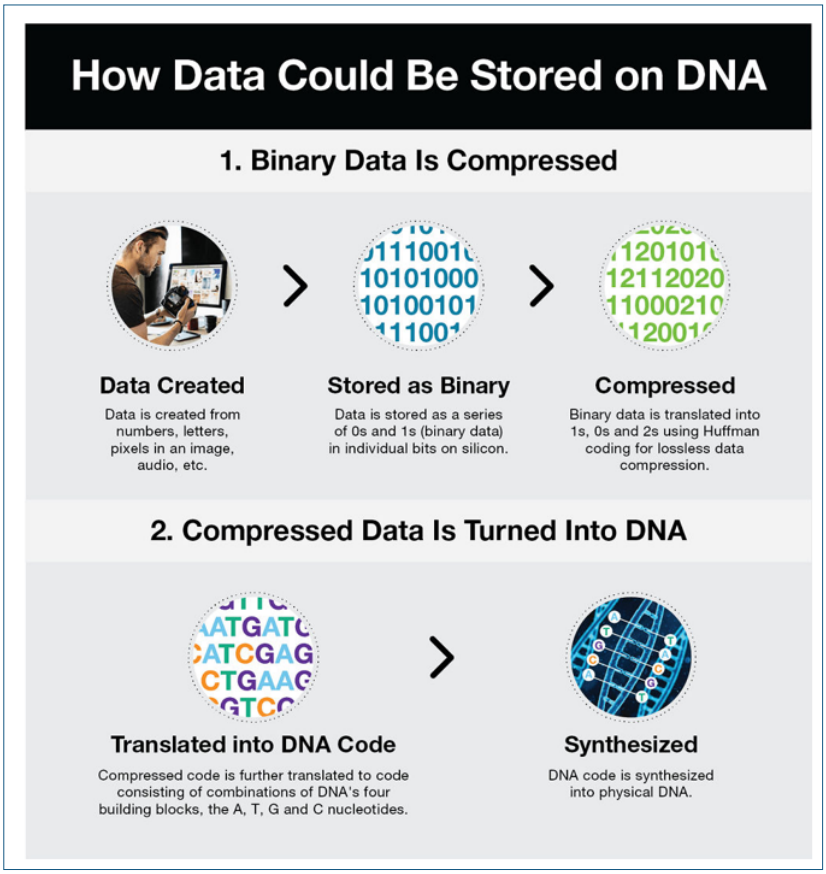

Figure 5: Overview of the Research Findings 3.

Table 1: An Overview of Available Data Materials.

Company Name, Country of Origin and Launch Date	Main Features of the Technology	Selected markets
Applied DNA Sciences, United States, 1983	Botanical DNA fragments, detection by PCR and CE, an encapsulation system	Product authentication, supply chain traceability, brand protection, anti-counterfeiting, textiles, pharmaceuticals, etc.
Haelixa, Switzerland, 2016	Synthetic DNA tags, detection by PCR, DNA enclosed in silica	Product authentication, supply chain traceability, intellectual property protection, etc.
Selectamark Security Systems (SelectaDNA), United Kingdom, 1986	Laboratory analysis of DNA for owner identification if microdots are absent (DNA serves as an alternative authentication solution)	Asset protection and recovery, securing high-value items, art and jewelry authentication, IT equipment and vehicle security, forensic applications, theft prevention and deterrence, etc.
TraceTag (CypherMark), United Kingdom / Norway, 2001	Synthetic DNA with unique primers, authorized access to primer sequences, detection using qPCR	Brand safeguarding, industrial applications, cash security, security of documentation, oil and fuel tracking, anti-counterfeiting measures, etc.
Holoptica, United States, 2012	Synthetics DNA tags (100 nucleotides), integration with inkjet cartridges	Artwork, documents and assets protection, verifying product authenticity, food tracking, etc.
DNA Technology, United States, 1993	DNA-laced ink, combination of DNA synthetic segments and optical taggants	Memorabilia and collectibles, limited edition artwork, pharmaceuticals, apparel and luxury goods, health and beauty industry, etc.
Tagsmart, United Kingdom, 2015	Synthetic DNA tags, secure Certificate of Authenticity	Artwork, securing collectibles, verifying paper documents, book manufacturing, etc.

DNA Guardian, Australia, 2007	UV-detectable stain, detection using pyrosequencing	Asset marking, crime prevention, artwork protection, theft deterrence, etc.
Aanika Biosciences, United States, 2018	Genetically modified Bacillus subtilis as an encapsulation system for DNA tag	Agriculture and food production, textiles, etc.

Discussions and Future Directions

The commercial trajectory of DNA data storage has moved beyond conceptual validation into the early stages of industrial deployment. High-profile initiatives—such as the 2026 roadmap by **Atlas Data Storage** (a Twist Bioscience spin-off) targeting terabyte-scale archival—demonstrate a shift toward practical, high-throughput systems.

The Hot vs. Cold Storage Dichotomy

Currently, DNA-based systems are optimized for **"Glacial" or "Cold" storage**—data that requires extreme longevity but infrequent access. While "warm" or "hot" storage (frequent access) remains a challenge due to synthesis latency, the development of **enzymatic DNA synthesis (EDS)** is the primary catalyst for change.

Unlike traditional phosphoramidite chemistry, EDS platforms (developed by leaders like **DNA Script**) offer faster, aqueous-based assembly, which is essential for scaling to the terabyte and petabyte levels within the next decade.

Economic Scalability and Industrial Adoption

The "cost-per-bit" is expected to decline sharply as automated, massively parallel synthesis chips reach the market. For organizations, the investment in DNA storage is a hedge against "bit rot" and the constant need for data migration between fragile electronic media.

- **Early Adopters:** Government archives, media conglomerates, and financial institutions are transitioning to molecular media to secure high-value records for centuries rather than decades.
- **Metadata and Retrieval:** Future success hinges on **random-access capabilities**, where specific files can be retrieved using molecular "indices" or CRISPR-based search tools without sequencing the entire library.

Environmental Sustainability and ESG Goals

As data centers face increasing scrutiny for their massive energy consumption and carbon footprints, DNA storage provides a "green" alternative.

- **Zero-Power Maintenance:** Once encoded, DNA requires no electricity for preservation, unlike the cooling-intensive server farms of today.
- **Waste Reduction:** Molecular storage drastically reduces e-waste, as the storage medium is a biological polymer rather than heavy metals and rare-earth elements found in silicon hardware.

"Toward a Bio-Digital Convergence"

The fusion of **Artificial Intelligence** and **Synthetic Biology** is creating a new era of data management. By 2026, we anticipate the first commercial "cloud-integrated" DNA tiers, where users can move data into a molecular archive as easily as they use traditional cloud backups. While technical hurdles in synthesis speed remain, a phased investment approach allows organizations to lead in technological prowess. Moving toward a **"DNA-of-Things"** paradigm—where every physical object can carry its own digital identity—will redefine our relationship with information, marking a permanent shift in how humanity preserves its digital legacy.

Conclusions

DNA-based data storage has transitioned from a theoretical curiosity to a viable, high-density archival solution capable of addressing the global "zettabyte crisis." Our analysis confirms that while the biological medium offers unparalleled longevity and a theoretical density of **1 exabyte per gram**, its immediate utility is defined by its role within a tiered storage hierarchy.

The "Glacial" Archival Tier

This research underscores that DNA is not a replacement for silicon-based hot storage (RAM or SSDs) but rather a revolutionary successor to magnetic tape. Its primary strength lies in **"Glacial Storage"**—environments where data must remain immutable and power-independent for centuries. The current 2026 benchmarks indicate that while retrieval latencies remain in the hour-to-day range, the **total cost of ownership** (TCO) over a 50-year horizon becomes increasingly competitive due to the elimination of data migration and energy-intensive cooling requirements.

Overcoming the Scaling Bottleneck

The "competitiveness gap" is currently tethered to the throughput of phosphoramidite synthesis. However, the findings suggest a pivotal shift toward **enzymatic assembly** and **combinatorial synthesis**. These second-generation "writing" technologies, bolstered by AI-driven error-correction and predictive sequence optimization, are essential for scaling from megabyte-scale pilots to the **terabyte-per-day** throughput required for enterprise adoption.

Integration of Artificial Intelligence (AI)

The role of AI has evolved from simple encoding to a fundamental component of the **"Molecular Operating System."** Machine learning algorithms now mitigate the stochastic noise inherent in biochemical processes, enabling high-fidelity data recovery even from partially degraded samples. This synergy between informatics and biotechnology is what will ultimately drive the commoditization of DNA archives.

Final Outlook

Ultimately, the successful integration of DNA into the global data infrastructure requires a multidisciplinary commitment to **standardization**. Establishing unified "molecular codecs" and automated "DNA-to-Digital" interfaces is the final hurdle. As we look toward 2030, DNA storage stands as a sustainable, secure, and physically compact solution that ensures our civilization's digital legacy is no longer at the mercy of fragile hardware, but is instead preserved in the resilient code of life itself.

Supplementary Information

The various original data sources some of which are not all publicly available, because they contain various types of private information. The available platform provided data sources that support the exploration findings and information of the research investigations are referenced where appropriate.

Acknowledgments
The authors would like to acknowledge GOOGLE Deep Mind Research with its associated pre-prints access platforms. This research was deployed and utilized under the platform provided by GOOGLE Deep Mind which is under the support of the GOOGLE Research and the GOOGLE Research Publications under GOOGLE Gemini platform. ChatGPT was used towards checking for any grammatical errors within the manuscript.

Declarations

- Funding
  No Funding was provided for the conduction concerning this research.
- Conflict of interest/Competing interests
  There are no Conflict of Interest or any type of Competing Interests for this research.
- Ethics approval
  The authors declares no competing interests for this research.
- Consent to participate
  The authors have read, approved the manuscript and have agreed to its publication.
- Consent for publication
  The authors have read, approved the manuscript and have agreed to its publication.
- Availability of data and materials
  The various original data sources some of which are not all publicly available, because they contain various types of private information. The available platform provided data sources that support the exploration findings and information of the research investigations are referenced where appropriate.
- Code availability
  Mentioned in details within the Acknowledgements section.
- Authors' contributions
  Described in details within the Acknowledgements section.

References

1. Zarif Bin Akhtar. Artificial intelligence within medical diagnostics: A multi-disease perspective. Artificial Intelligence in Health 5173. https://doi.org/10.36922/aih.5173
2. Akhtar, Z. B., & Rozario, V. S. (2025). AI Perspectives within Computational Neuroscience: EEG Integrations and the Human Brain. Artificial Intelligence and Applications. https://doi.org/10.47852/bonviewaia52024174
3. Zarif Bin Akhtar. (2025). Beyond Perception: A Comprehensive Investigation into the Advancements, Challenges & Ethical Dimensions of AI and Computer Vision. Real-World AI Systems, 1(1), 1–27. https://doi.org/10.30564/rwas.v1i1.9577
4. Doricchi, A., Platnich, C. M., Gimpel, A., Horn, F., Earle, M., Lanzavecchia, G., ... & Garoli, D. (2022). Emerging approaches to DNA data storage: challenges and prospects. ACS nano, 16(11), 17552-17571.
5. Wang, S., Mao, X., Wang, F., Zuo, X., & Fan, C. (2024). Data storage using DNA. Advanced Materials, 36(6), 2307499.
6. Hao, Y., Li, Q., Fan, C., & Wang, F. (2021). Data storage based on DNA. Small Structures, 2(2), 2000046.
7. Buko, T., Tuczko, N., & Ishikawa, T. (2023). DNA data storage. BioTech, 12(2), 44.
8. Matange, K., Tuck, J. M., & Keung, A. J. (2021). DNA stability: a central design consideration for DNA data storage systems. Nature communications, 12(1), 1358.
9. Akash, A., Bencurova, E., & Dandekar, T. (2024). How to make DNA data storage more applicable. Trends in biotechnology, 42(1), 17-30.
10. Lim, C. K., Nirantar, S., Yew, W. S., & Poh, C. L. (2021). Novel modalities in DNA data storage. Trends in biotechnology, 39(10), 990-1003.
11. Chen, Y. J., Takahashi, C. N., Organick, L., Bee, C., Ang, S. D., Weiss, P., ... & Strauss, K. (2020). Quantifying molecular bias in DNA data storage. Nature communications, 11(1), 3264.
12. Shomorony, I., & Heckel, R. (2022). Information-theoretic foundations of DNA data storage. Foundations and Trends® in Communications and Information Theory, 19(1), 1-106.
13. Yu, M., Tang, X., Li, Z., Wang, W., Wang, S., Li, M., ... & Chen, C. (2024). High-throughput DNA synthesis for data storage. Chemical Society Reviews, 53(9), 4463-4489.
14. Raza, M. H., Desai, S., Aravamudhan, S., & Zadegan, R. (2023). An outlook on the current challenges and opportunities in DNA data storage. Biotechnology advances, 66, 108155.
15. Coudy, D., Colotte, M., Luis, A., Tuffet, S., & Bonnet, J. (2021). Long term conservation of DNA at ambient temperature. Implications for DNA data storage. PLoS One, 16(11), e0259868.
16. Yang, S., Boegels, B. W., Wang, F., Xu, C., Dou, H., Mann, S., ... & de Greef, T. F. (2024). DNA as a universal chemical substrate for computing and data storage. Nature Reviews Chemistry, 8(3), 179-194.
17. Antkowiak, P. L., Lietard, J., Darestani, M. Z., Somoza, M. M., Stark, W. J., Heckel, R., & Grass, R. N. (2020). Low cost DNA data storage using photolithographic synthesis and advanced information reconstruction and error correction. Nature communications, 11(1), 5345.
18. Organick, L., Nguyen, B. H., McAmis, R., Chen, W. D., Kohll, A. X., Ang, S. D., ... & Strauss, K. (2021). An empirical comparison of preservation methods for synthetic DNA data storage. Small Methods, 5(5), 2001094.
19. Gimpel, A. L., Stark, W. J., Heckel, R., & Grass, R. N. (2023). A digital twin for DNA data storage based on comprehensive quantification of errors and biases. Nature Communications, 14(1), 6026.
20. Nguyen, B. H., Takahashi, C. N., Gupta, G., Smith, J. A., Rouse, R., Berndt, P., ... & Strauss, K. (2021). Scaling DNA data storage with nanoscale electrode wells. Science advances, 7(48), eabi6714.
21. Xu, C., Zhao, C., Ma, B., & Liu, H. (2021). Uncertainties in synthetic DNA-based data storage. Nucleic acids research, 49(10), 5451-5469.
22. Ezekannagha, C., Becker, A., Heider, D., & Hattab, G. (2022). Design considerations for advancing data storage with synthetic DNA for long-term archiving. Materials Today Bio, 15, 100306.
23. Kim, S. J., Jung, W. B., Jung, H. S., Lee, M. H., Heo, J., Horgan, A., ... & Ham, D. (2023). The bottom of the memory hierarchy: Semiconductor and DNA data storage. MRS bulletin, 48(5), 547-559.
24. Dong, Y., Sun, F., Ping, Z., Ouyang, Q., & Qian, L. (2020). DNA storage: research landscape and future prospects. National Science Review, 7(6), 1092-1107.
25. Bee, C., Chen, Y. J., Queen, M., Ward, D., Liu, X., Organick, L., ... & Ceze, L. (2021). Molecular-level similarity search brings computing to DNA data storage. Nature communications, 12(1), 4764.
26. Antkowiak, P. L., Koch, J., Nguyen, B. H., Stark, W. J.,

Strauss, K., Ceze, L., & Grass, R. N. (2022). Integrating DNA encapsulates and digital microfluidics for automated data storage in DNA. Small, 18(15), 2107381.
27. Zhang, C., Wu, R., Sun, F., Lin, Y., Liang, Y., Teng, J., ... & Yan, H. (2024). Parallel molecular data storage by printing epigenetic bits on DNA. Nature, 634(8035), 824-832.
28. Lenz, A., Siegel, P. H., Wachter-Zeh, A., & Yaakohi, E. (2020, May). Achieving the capacity of the DNA storage channel. In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 8846-8850). IEEE.
29. Chen, K., Zhu, J., Boskovic, F., & Keyser, U. F. (2020). Nanopore-based DNA hard drives for rewritable and secure data storage. Nano letters, 20(5), 3754-3760.
30. Lin, K. N., Volkel, K., Tuck, J. M., & Keung, A. J. (2020). Dynamic and scalable DNA-based information storage. Nature communications, 11(1), 2981.
31. Grass, R. N., Heckel, R., Dessimoz, C., & Stark, W. J. (2020). Genomic encryption of digital data stored in synthetic DNA. Angewandte Chemie International Edition, 59(22), 8476-8480.
32. Meiser, L. C., Antkowiak, P. L., Koch, J., Chen, W. D., Kohll, A. X., Stark, W. J., ... & Grass, R. N. (2020). Reading and writing digital data in DNA. Nature protocols, 15(1), 86-101.
33. Song, L., Geng, F., Gong, Z. Y., Chen, X., Tang, J., Gong, C., ... & Yuan, Y. J. (2022). Robust data storage in DNA by de Bruijn graph-based de novo strand assembly. Nature communications, 13(1), 5361.