ICDS 2026

The Twentieth International Conference on Digital Society

ISBN: 978-1-68558-382-8

May 24 - 28, 2026

Venice, Italy

ICDS 2026 Editors

Lasse Berntzen, University of South-Eastern Norway, Norway

Olga Levina, Technische Hochschule Brandenburg, Germany

ICDS 2026

Forward

The Twentieth International Conference on Digital Society (ICDS 2026), held between May 24, 2026, and May 28, 2026, in Venice, Italy, continued a series of events covering a large spectrum of topics related to advanced networking, applications, and system technologies in a digital society. Nowadays, most economic activities and business models are driven by the unprecedented evolution of theories and technologies. The reflection of these achievements into our society is present everywhere, and it is only question of user education and business models optimization towards a digital society.

Progress in cognitive science, knowledge acquisition, representation, and processing helped to deal with imprecise, uncertain, or incomplete information. Management of geographical and temporal information becomes a challenge, in terms of volume, speed, semantic, decision, and delivery. Information technologies allow optimization in searching and interpreting data, yet special constraints imposed by the digital society require on-demand, ethics, and legal aspects, as well as user privacy and safety.

The variety of the systems and applications and the heterogeneous nature of information and knowledge representation require special technologies to capture, manage, preserve, interpret, and deliver the content and documents related to a particular target.

We take here the opportunity to warmly thank all the members of the ICDS 2026 technical program committee, as well as all the reviewers. The creation of such a high-quality conference program would not have been possible without their involvement. We also kindly thank the authors who dedicated time and effort to contribute to ICDS 2026. We truly believe that, thanks to all these efforts, the final conference program consisted of top-quality contributions. We also thank the members of the ICDS 2026 organizing committee for their help in handling the logistics of this event.

We hope that ICDS 2026 was a successful international forum for the exchange of ideas and results between academia and industry for the promotion of progress in a digital society.

ICDS 2026 Chairs

ICDS 2026 Steering Committee
Lasse Berntzen, University of South-Eastern Norway, Norway
Claus-Peter Rückemann, Universität Münster / DIMF / Leibniz Universität Hannover, Germany
Theo Lynn, Irish Institute of Digital Business, Dublin City University, Ireland
Olga Levina, Technische Hochschule Brandenburg, Germany
Claudia Heß, IU Internationale Hochschule, Germany

ICDS 2026 Publicity Chairs
José Miguel Jiménez, Universitat Politecnica de Valencia, Spain
Francisco Javier Díaz Blasco, Universitat Politècnica de València, Spain
Ali Ahmad, Universitat Politècnica de València, Spain
Sandra Viciano Tudela, Universitat Politecnica de Valencia, Spain
Laura Garcia, Universidad Politécnica de Cartagena, Spain

ICDS 2026
Committee

ICDS 2026 Steering Committee

Lasse Berntzen, University of South-Eastern Norway, Norway
Claus-Peter Rückemann, Universität Münster / DIMF / Leibniz Universität Hannover, Germany
Theo Lynn, Irish Institute of Digital Business, Dublin City University, Ireland
Olga Levina, Technische Hochschule Brandenburg, Germany
Claudia Heß, IU Internationale Hochschule, Germany

ICDS 2026 Publicity Chairs

José Miguel Jiménez, Universitat Politecnica de Valencia, Spain
Francisco Javier Díaz Blasco, Universitat Politècnica de València, Spain
Ali Ahmad, Universitat Politècnica de València, Spain
Sandra Viciano Tudela, Universitat Politecnica de Valencia, Spain
Laura Garcia, Universidad Politécnica de Cartagena, Spain

ICDS 2026 Technical Program Committee

Chiniah Aatish, University of Mauritius, Mauritius
Iván Abellán, University of Luxembourg, Luxembourg
Laura Alcaide Muñoz, University of Granada, Spain
Ludivine Allienne, Université Picardie Jules Verne - laboratoire CURAPP-ESS, France
Subia Ansari, Purdue University, West Lafayette, USA
Kambiz Badie, ICT Research Institute, Iran
Alessandra Bagnato, Softeam, France
Ilija Basicevic, University of Novi Sad, Serbia
Najib Belkhayat, Cadi Ayyad University of Marrakech, Morocco
Lasse Berntzen, University of South-Eastern Norway, Norway
Aljosa Jerman Blazic, SETCCE Ltd. / IT association at Chamber of commerce, Slovenia
Mahmoud Brahimi, University of Msila, Algeria
Justin F. Brunelle, The MITRE Corporation, USA
Erik Buchmann, Universität Leipzig / ScaDS.AI, Germany
Marcos F. Caetano, University of Brasília, Brazil
Maria Chiara Caschera, CNR-IRPPS, Italy
Bidisha Chaudhuri, University of Amsterdam, The Netherlands
Sunil Choenni, Dutch Ministry of Justice and Security / Rotterdam University of Applied Sciences, Netherlands
Yul Chu, University of Texas Rio Grande Valley (UTRGV), USA
Andrei V. Chugunov, ITMO University, St. Petersburg, Russia
María E. Cortés-Cediel, Universidad Complutense de Madrid, Spain
Vladimir Costas-Jauregui, Universidad Mayor de San Simón, Bolivia
Arthur Csetenyi, Budapest Corvinus University, Hungary
Ibibia K. Dabipi, University of Maryland Eastern Shore, USA

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

Ownership and Flow Primitives for Scalable Consent Management in Digital Public Infrastructures

Rohith Vaidyanathan, Dev Shinde, Praseeda, Srinath Srinivasa
Web Science Lab
International Institute of Information Technology, Bangalore
Bengaluru, Karnataka, India
e-mail: {rohith.vaidyanathan, shindedev.hemravi, praseeda, sri}@iiitb.ac.in

***Abstract*—Digital Public Infrastructures (DPIs) represent networks of open technology standards, applications, services, and digital assets made available for the public good. One of the key challenges in DPI design is to resolve complex issues of consent, scaled over large populations. While the primary objective of consent management is to empower the data owner, ownership itself can come with variegated morphological forms with different implications over consent. This paper addresses the question of representing modes of ownership of digital assets and their corresponding implications for consensual data flows in a DPI. It proposes a set of foundational abstractions to represent them, incorporating a formalised data ownership model that enables end-to-end traceability of consent, fine-grained flow control over data sharing, and alignment with evolving legal and regulatory frameworks.**

Keywords-Digital Public Infrastructure; Consent Management; e-Government; Data Ownership; Data Flow Control; Privacy Architecture; Data Governance.

I. INTRODUCTION

Digital Public Infrastructures (DPIs) [1] represent open ecosystems of digital services, applications and assets made available for public good. DPI services like digital identity, lockers, catalogs, wallets and payment infrastructures have streamlined several e-governance activities and economic, legal and social transactions [2]. There have been several initiatives in the past where digital assets and services were designed to be owned by and available for the public, such as free software movements and creative commons. However, the term DPI refers to systemic, scalable infrastructure that is meant to act as a "digital backbone" for the entire society, enabling access to both public and private services like healthcare, education, and financial inclusion. The role of the state and institutional players is central to DPI design, distinguishing it from community-driven approaches.

As public interactions become more digital, vast amounts of sensitive data are exchanged through DPIs. Without proper mechanisms, data owners may lose control over how their data is used post-exchange. Various countries have implemented data protection and consent management laws to address this [3]. Consent management is modeled as an interception of a data flow in a DPI, to ensure that the flow is consistent with the consent of the data owner.

In its simplest form, a consent architecture elicits explicit consent from the Data Owner (DO) in response to a data request, following the Autonomous Authorization (AA) model [4]. However, as DPI implementations scale, the AA architecture becomes insufficient due to consent desensitization from frequent requests. Additionally, if data is shared by an independent custodian on behalf of the owner, the custodian needs data sharing policies set by the owner. As a result, *policy-based* consent management architectures become necessary for DPI implementations [5].

Consent management also has several other nuances. Consent is closely related to ownership, which can have several morphological forms. Ownership may be *delegated* or *pledged*, and partial ownership or privileges can be *conferred* all with implications on how consent is granted and enforced. Consent management also extends beyond access control: consent provisions may need enforcement even after access is granted, including whether data may be shared further, aggregated with other data, the purposes for which they may be used, and requirements of suitable notifications to the data owner.

While DPI provides the primary motivating context for this work, the ownership and flow primitives proposed here are designed to be architecture-agnostic. The same abstractions apply equally to enterprise data governance, academic credential ecosystems, cross-institutional healthcare networks, and any open-ended data exchange environment where ownership semantics and post-access governance are required.

This paper proposes an architecture that addresses complex forms of ownership and flow primitives required for supporting consensual data flows.

The remainder of this paper is organised as follows. In Section II, we survey related literature on data sharing, access control, and consent management. In Section III, we introduce the formal model for public data flows. In Section IV, we describe the consent flow architecture, including connections and X-nodes. In Section V, we define data exchange operations. In Section VI, we analyse adversarial scenarios and the system's resilience to them. Finally, Section VII concludes the paper and outlines future work.

II. RELATED LITERATURE

Although the term DPI is a relatively recent introduction, some of its core data-related challenges, such as inter-organizational and open-ended sharing of sensitive data, have been addressed for several years [6]–[9]. Inter-organizational data sharing addresses workflows spanning multiple organizations and governance structures [10][11].

 ISBN: 978-1-68558-382-8

When access requests cross organizational boundaries, they could potentially lead to unsafe data access. Innovations in this space include extensions of Role-Based Access Control (RBAC) [12][13], attribute-based models [14][15], or hybrid models [16][17]. However, these access control mechanisms focus on the authorization decision whether to grant access but do not address post-access governance: what recipients can do with data after receiving it, how ownership evolves through sharing, or how to revoke access after cross-organizational transfer.

Consent management in public data flows goes beyond access control [5]. Research has addressed philosophical issues of what makes consent meaningful [4][18], as well as various consent architectures [5][19][20]. Existing systems like DI-CON (Domain Independent Consent Management) [21] and CMA (Consent Management Architecture) [22] treat consent primarily as an authorization checkpoint rather than dynamic user-centric control. Most existing consent managers employ AA [4], where explicit consent is elicited for each request. However, AA is not scalable for DPIs due to: (i) consent fatigue from frequent requests leading to desensitization; (ii) inability to handle custodial sharing where data is managed on behalf of owners; (iii) infeasibility of human intervention for institutional data owners; and (iv) request volumes in large-scale DPI implementations that make individual approval impractical. Policy-based consent [5] addresses these limitations by enforcing pre-defined rules in a domain-agnostic manner. Questions persist about what data flows one can legitimately control by virtue of ownership [23][24].

Regulations in India [25] and Europe [26] have outlined mechanisms for lawful consent. Pioneering frameworks include X-road [27][28]. As data is shared in an open-ended fashion in DPIs, data ownership morphs into different constructs requiring meaningful modeling regarding their impact on consent.

An effective consent management system must encompass not only access control but also explicitly address issues of ownership to enable ongoing user control over data following its sharing and ensure compliance with evolving data policies. In this paper, we propose a comprehensive model that systematically addresses all four of these critical dimensions of consent management.

Table I defines the core terminology used throughout this paper.

TABLE I. KEY TERMINOLOGY.

Term	Definition
DO	Data Owner: Controls data resource with full authority
DR	Data Requester: Seeks access to data owned by another
DS	Data Subject: Entity described by the data

III. MODELING PUBLIC DATA FLOWS

Any public infrastructure poses a complex interplay between individual rights and public interest. We model a public data flow network as semantic containers representing ownership boundaries of stakeholders, as defined in Equation 1:

$$DPI = (A, L, C) \quad (1)$$

Here, A is a set of *agents* or *stakeholders* who assert ownership and play roles like Data Owner (DO), Data Requester (DR), or Data Subject (DS). L is a set of containers called *Access Policy Domains (APD)* or *lockers*, representing semantic boundaries where data ownership is enforced. $C \subseteq L \times L$ represent data flow pipelines called *Connections*, established
between lockers as legitimate pathways for data exchange, made legitimate by underlying contracts encapsulating data sharing policies and applicable regulations.

An *artifact* is a logical unit of data subject to ownership and consent that flows through connections. It represents a data resource (e.g., a document), though a single resource may have multiple artifacts. The consent service regulates artifact storage and flow, while a separate resource service manages actual resources.

IV. CONSENT FLOW ARCHITECTURE

We illustrate consent flow using a running example: A student s uses a DPI to obtain her degree from university u and applies for a job with company c.

Degree granting: Student s requests transcripts from university u. While s is the owner of her transcripts, she cannot modify them unilaterally, she is the *conferred owner*, while u remains the *primary owner*.

Job application: Student s shares her transcripts and degree credentials with company c. As part of this share, the company only requires *access rights*, not ownership. "Sharing" here means granting access rights.

Credentials Verification: Company c seeks verification from university u. The university may verify without consent of s (as primary owner) unless regulations require otherwise.

Job offer: Company c requires s to *pledge* her certificates as collateral as long as she is in full-time employment with them. As a pledged asset, s retains access but cannot transfer or re-pledge them.

Job Contract: Upon pledging, company c becomes the *pledged owner* with limited rights, while university u remains the *primary owner*. Student s continues as *conferred owner* in a constrained manner.

Contract Termination: Once the contract ends, transcripts are returned to s, company c no longer has access, and s is no longer subject to pledge restrictions. The pledge is *reverted*.

A. Connection between Lockers

Before data transactions, agents establish a *connection* between lockers, formalizing the terms of a consensual transaction. A DO publishes *connection endpoints* representing connection terms. A DR can connect any of their lockers to a given endpoint (Figure 1).

A *connection type* specifies a schema representing terms and conditions for connection establishment, encoding rules from the DO's policy and applicable regulations. Rules are

expressed as Event-Condition-Modality-Action (ECMA) statements [5], where each rule binds a triggering *event* (e.g., a data access request) to a *condition* (e.g., stated purpose, requester identity) and assigns a normative *modality* Obligated (O), Permitted (P), or Forbidden (F) over a resulting *action*. These OPF modalities encode the normative intent of a policy rather than serving as mere technical flags. *Obligated* (O) denotes that an action *must* occur; the requester cannot proceed until it is fulfilled (e.g., submitting ethical clearance before accessing sensitive health data). *Permitted* (P) denotes that an action *may* occur; it represents a green light that can be further restricted by layered policies. *Forbidden* (F) denotes that an action *must not* occur. Together, OPF move consent management beyond binary allow/deny access control into a richer normative space where every artifact operation carries an enforceable legal and ethical standing. A sample ECMA rule takes the form:

```
ON [RequestAccess] IF [purpose =
"verification"] THEN PERMIT [read]
WITH [validity = 7 days, share =
false]
```

Multiple rules from different institutional sources may fire simultaneously on the same artifact. When their modalities are complementary, for example when a regulation permits an action and an organizational policy further restricts it, they compose predictably: the stricter modality prevails. When modalities directly conflict (e.g., one source requires an action that another forbids), the system flags an irreconcilable policy exception rather than silently resolving it. Such conflicts are escalated to a designated human agent for resolution, reflecting the principle that normative contradictions between institutions require human judgment, not automated override. Depending on the deployment context, the policy evaluation layer may be realized as a rule-based engine, a logic interpreter, or an AI-assisted policy reasoner; the ECMA structure is substrate-agnostic. Readers are directed to [5] for a full formal treatment of the four-layer architecture, normative axioms, and conflict resolution semantics.

The connection lifecycle (Figure 2) begins with the PUBLISHED state when the DO publishes endpoints. When the DR selects an endpoint, it transitions to ESTABLISHED. After fulfilling obligations, it becomes LIVE, enabling the data exchange operations detailed in Section V. Either party may REVOKE during ESTABLISHED or LIVE states. After completion or expiry, connections are CLOSED.

B. X-nodes

Once a connection is live, *artifacts* representing data resources flow through the connections. Note that the data resource itself does not flow through a connection; only the artifacts that represent privileges over data resources are exchanged through these connections.

We propose a formal data structure called *X-nodes* for consent artifacts that encapsulate terms of consent and enforce post-conditions. Table II summarises the essential fields and operations for each X-node type.

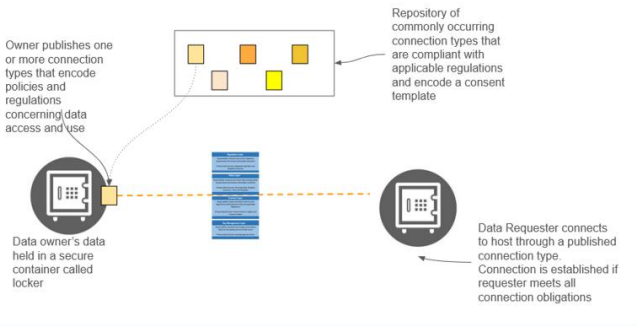


Figure 1. Establishing a consensual pathway governed by the four-layer architecture of consent [5].

X-nodes are of three types. The i-node ("information" node) represents the primary location of a resource with full authority for the primary owner. The s-node ("shadow" node) is received when data is conferred, allowing read/write access as per the policy specified by the DO with a pointer to the original artifact. The v-node ("virtual" node) represents an access privilege granted for a resource, containing only a pointer to the original artifact with a mandatory validity period.

Three critical ownership fields define the control structure of each X-node: `creator` identifies the agent who originally instantiated the X-node. `primary_owner` (PO) represents the entity with fundamental authority over the resource (e.g., government for driving licenses, university for degree certificates), which can confer, modify, or revoke the resource. `current_owner` (CO) indicates the entity presently in possession of the X-node with operational control. When `primary_owner` and `current_owner` differ, the X-node is said to be *locked*, preventing transfer, conferment, or pledging until the constraint is resolved. Each X-node has additional essential fields (Table II): `shadows_list`, `v-node_list`, `pointer_to_resource`, `pointer_to_original`, `validity`, `purpose`, and `provenance`.

Essential post-conditions specify actions the Data Requester can perform: `transfer`, `confer`, `share`, `collateral`, `subset`, and `download`.

V. Data Exchange Operations

The following kinds of data exchange operations are defined, each with different semantics over ownership and consent:

SHARE: The DO shares an *access privilege* to resource *r*. A v-node sent to the data requester is formally of the form $dr.v(do.i(r))$ or $dr.v(do.s(r))$, where dr is the data requester and do is the data owner. The v-node represents a pointer to a corresponding i-node or s-node, which in turn represents data resource *r*. This v-node can be further shared to yet another requester, creating an "access tunnel" of the form $dr2.v(dr.v(do.i(r)))$. The actual access to resource *r* traverses the tunnel and is initiated by the last i-node or s-node. The resource server receives the *access tunnel*, approving access

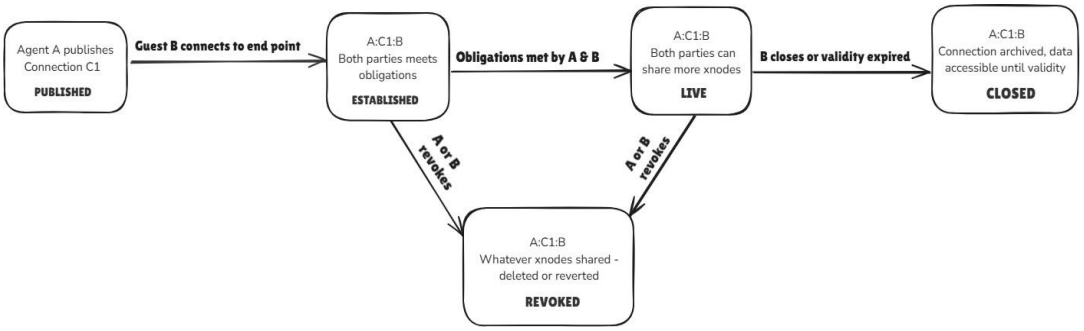


Figure 2. Connection lifecycle illustrating Agent B establishing a connection of type C1 published by Agent A.

TABLE II. ESSENTIAL FIELDS AND OPERATIONS OF X-NODES WITH PRACTICAL EXAMPLES.

X-node	Function	Essential fields	Essential post-conditions	Example Use Case
i-node	Primary location of resource	`creator`, `primary_owner`, `current_owner`, `shadows_list`, `v-node_list`, `pointer_to_resource`, `purpose`, `provenance`	`transfer`, `confer`, `share`, `collateral`, `subset`, `download`	Government holds i-node of citizen's driving license with full authority to issue, modify, suspend, or revoke
v-node	Access privilege to an i-node	`creator`, `current_owner`, `pointer_to_original`, `validity`, `v-node_list`, `purpose`, `provenance`	`transfer`, `share`, `download`	Car rental company receives temporary v-node to verify driver's license validity for 7 days, without owning the license
s-node	Conferred or pledged ownership	`creator`, `primary_owner`, `current_owner`, `pointer_to_original`, `shadows_list`, `v-node_list`, `pointer_to_resource`, `purpose`, `provenance`	`transfer`, `share`, `collateral`, `subset`, `download`	Citizen receives s-node as conferred owner of driving license, can present it but cannot modify; or pledges it as ID proof

only if *do* is `primary_owner`. Since access comes through the owner's i-node or s-node, the DO is aware of every access (Figure 3(ii)). *Example:* A car rental company requests to verify a customer's driving license. The customer (DO) creates a v-node for the conferred license and sets a validity of (say) 7 days, allowing the rental company (DR) temporary read-only access to verify license validity without transferring ownership or providing a permanent copy.

CONFER: The DO shares an immutable copy to the DR, who becomes owner (e.g., certificates, licenses, tickets). The DO creates an s-node that is sent to DR. The s-node's `primary_owner` and `current_owner` are set to DR, while the i-node's `primary_owner` remains DO and `current_owner` becomes DR, making it *locked* (Fig-ure 3(iii)). Conferred owners access resources read-only through s-nodes. *Example:* A university issues a degree certificate to a student. The university (DO) creates an s-node from the i-node it owns and confers it to the student (DR). The student can present the certificate to employers but cannot modify its contents. If the university needs to correct an error (e.g., a misspelled name), only the university can make changes to the original i-node.

TRANSFER: Here, ownership *transfers* from DO to DR. The X-node itself transfers to DR with both PO and CO set to DR, providing complete ownership. The former DO loses access (Figure 3(i)). Transfer invalidates v-nodes and s-nodes pointing to it. Any forbidden post-condition set by the artifact creator continues across transfers. *Example:* A person sells their car registration documents to the new owner. The original owner (DO) transfers the i-node to the buyer (DR), who becomes both primary and current owner. The seller loses all access to the documents, and the buyer can now modify, share, or transfer them independently, subject only to restrictions set by the original creator (e.g., the vehicle authority).

COLLATERAL: Here, the resource is pledged as collateral. The DO sends the i-node to DR, keeping `primary_owner` as DO and setting `current_owner` to DR. The DR creates an s-node with `primary_owner` as DR and `current_owner` as DO. Both X-nodes are locked, preventing further pledge, conferment, or transfer until released (Figure 3(iv)). *Example:* A job applicant pledges their degree certificate as part of an employment contract, agreeing not to use it for other applications during employment. The applicant (DO) sends the i-node to the employer (DR), who holds it as collateral. The employer issues an s-node back to the applicant, who retains the ability to present the certificate for verification but cannot pledge it elsewhere. Upon contract termination, the pledge is reverted and the i-node returns to the applicant.

Compositions: V-nodes from SHARE can be further shared (if permitted), creating *access tunnels* through multiple consent layers (Figure 4). V-nodes can be transferred, moving access pathways. S-nodes from conferment can be subject to SHARE, COLLATERAL, or TRANSFER (unless restricted). Locked artifacts from COLLATERAL cannot be transferred, conferred, or pledged, but SHARE is permitted.

CLOSE, REVOKE, REVERT: CLOSE ends an open connection after successful data sharing. Downloaded artifacts remain with DR; v-nodes continue until validity expires. REVOKE rolls back a connection, reverting i-nodes and s-nodes to original owners. REVOKE on artifacts deletes v-nodes (SHARE) or returns X-nodes (TRANSFER). REVERT reverses conferment or collateral even after connection closure, returning pledged i-nodes or deleting conferred s-nodes.

Together, these operations form a composable, policy-governed layer for consensual data exchange that supports fine-grained ownership tracking throughout the lifecycle of every artifact in the DPI.

VI. Adversarial Scenarios

In this section, we present a few adversarial scenarios and show how the proposed architecture addresses them.

Scenario 1: Impersonation. Bob creates a fake locker impersonating a legitimate vendor to access sensitive tender documents. Alice, a company (DO), publishes a connection endpoint for vendor registration. Bob (DR) connects his fraudulent locker, claiming to represent "XYZ Corp," attempting to access project specifications shared with verified vendors.

Handling: Alice's connection type specifies obligations requiring DR to share verifiable credentials (business registration, tax ID, digital attestations) before the connection becomes live. When Bob connects, the connection enters ESTABLISHED state but remains non-operational. Alice verifies the submitted credentials against authoritative sources (e.g., government registries). The connection transitions to LIVE only after successful verification. Since Bob cannot provide legitimate credentials, the connection remains in ESTABLISHED state, blocking all artifact exchange. Alice can REVOKE the connection if verification fails, ensuring data sharing occurs only with authenticated agents.

Scenario 2: External Replication. Alice (DO) shares confidential financial documents with an auditing firm (DR) via v-node with `download` set to false for compliance verification only. Bob, an employee with legitimate v-node access, attempts to circumvent controls by photographing the screen with his mobile device to leak information to competitors.

Handling: The client application enforces view-only constraints, disabling screenshots, screen recording, and downloads. However, the consent manager cannot prevent physical photography. To mitigate this, the resource displays dynamic watermarking embedding Bob's identity, timestamp, and access context, making unauthorized copies traceable. The consent manager logs comprehensive audit trails including Bob's identity, legal capacity (auditing firm employee), connection details, timestamps, and access purpose. If leaked documents are discovered, these logs and watermarks provide forensic evidence identifying Bob as the breach source, enabling Alice to pursue legal remedies.

Scenario 3: Cascaded Sharing. Alice (DO) shares medical records with Dr. Smith (DR1) via v-node. Dr. Smith attempts to cascade share to specialist Dr. Jones (DR2) for consultation without Alice's explicit consent, creating a v-node from his existing v-node.

Handling: If Alice prohibits re-sharing, the `share` post-condition is set to false, blocking Dr. Smith from cascading the share. If permitted, Dr. Smith creates a v-node for Dr. Jones, forming access tunnel $dr2.v(dr1.v(do.i(r)))$. All access requests traverse Alice's i-node, ensuring she is aware of all accesses. Alice can revoke either Dr. Smith's v-node (eliminating both accesses) or specifically Dr. Jones's cascaded v-node at any time, maintaining ultimate control over her medical records.

Scenario 4: Cross-Border. A DR is in a different legal jurisdiction from the DO. *Handling:* Cross-border data transfers are not currently supported within this model. The model assumes DR and DO are within the same jurisdictional boundary; extending it to cross-border scenarios is an important avenue for future work.

VII. Conclusion

In this paper, we propose an architecture addressing data security and governance as data transitions through its core states in a DPI: *at rest*, *in transit*, and *in use*.

The present work contributes a *semantic architecture*: a set of formally grounded ownership and flow primitives that serve as a technology-agnostic blueprint from which concrete systems can be derived. A working implementation of this architecture, the Anumati Consent Management System developed at the Web Science Lab, IIIT Bangalore, is publicly accessible at [29].

To summarise, we presented an architecture using X-nodes and Connections to support consensual, public data exchange that empowers data owners and supports compliance with diverse regulations, including personal and other data laws. A key strength of the framework is that its ownership and flow primitives are architecture-agnostic: they apply equally to enterprise data governance, academic credential networks, and healthcare interoperability. By managing data sharing through regulated pipelines called connections and by supporting detailed tracking of consent and ownership changes, our design allows data owners to retain control even after data is shared. Overall, implementing consent management at the DPI layer ensures that it can be both robust and adaptable, giving individuals genuine control and helping organizations remain compliant as data laws and expectations evolve.

References

[1] United Nations Development Programme, "Digital public infrastructure," Accessed: 2025-07-08, 2025. [Online]. Available: https : / / www . undp . org / digital / digital - public - infrastructure.

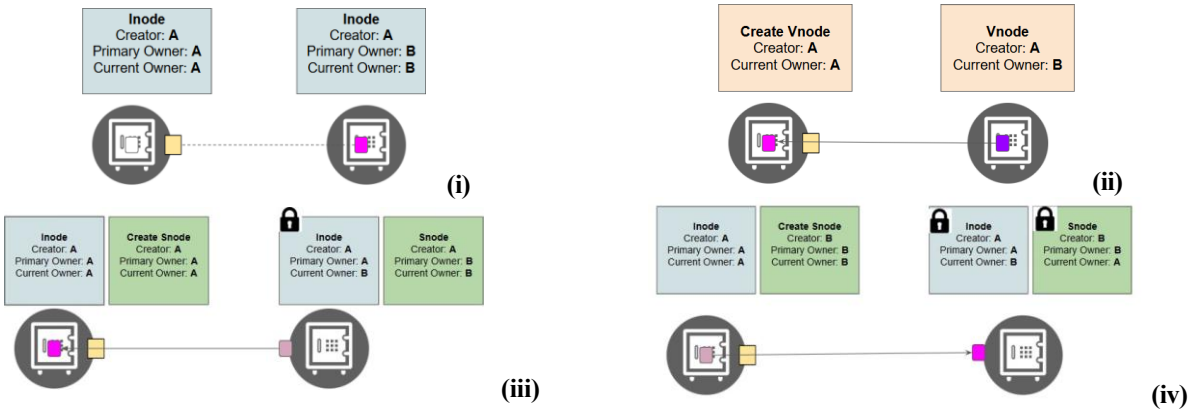


Figure 3. Sharing operators: (i) TRANSFER (ii) SHARE (iii) CONFER (iv) COLLATERAL.

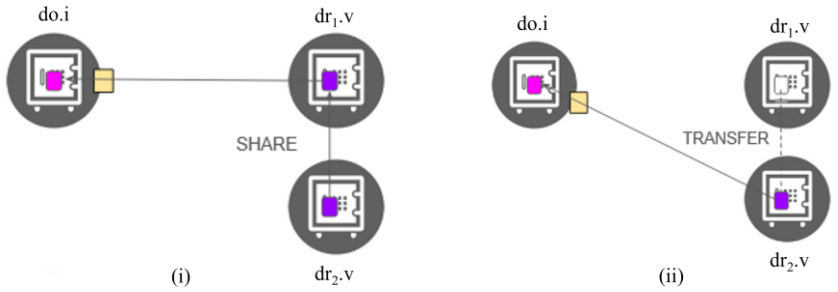


Figure 4. (i) Access tunnel for origin dr_2 to ground do after SHARE of SHARE. (ii) V-node ($dr_1.v$) transferred from dr_1 to dr_2.

[2] R. Bandura, M. McLean, and S. Sultan, "Unpacking the concept of digital public infrastructure and its importance for global development," *Center for Strategic and International Studies (CSIS)*, 2023.

[3] G. Gheorghiu et al., "The EU general data protection regulation implications for Romanian small and medium-sized enterprises," *Ovidius University Annals (Economic Sciences Series)*, vol. 18, no. 1, pp. 88–91, 2018.

[4] B. W. Schermer, B. Custers, and S. Van der Hof, "The crisis of consent: How stronger legal protection may lead to weaker consent in data protection," *Ethics and Information Technology*, vol. 16, no. 2, pp. 171–182, 2014.

[5] B. Ayyapane, R. Vaidyanathan, S. Srinivasa, S. Upadhyaya, and S. Vivek, "Consent service architecture for policy-based consent management in data trust," in *Proceedings of the 7th Joint International Conference on Data Science & Management of Data (11th ACM IKDD CODS and 29th COMAD)*, ACM, 2024.

[6] K. Houser and J. W. Bagby, "The data trust solution to data sharing problems," *Vanderbilt Journal of Entertainment & Technology Law*, 2023.

[7] S. Shrivastava and T. Srikanth, "A comprehensive consent management system for electronic health records in the healthcare ecosystem," in *Information Security and Privacy in Smart Devices: Tools, Methods, and Applications*, IGI Global, 2023, pp. 194–233.

[8] S. Stalla-Bourdillon, G. Thuermer, J. Walker, L. Carmichael, and E. Simperl, "Data protection by design: Building the foundations of trustworthy data sharing," *Data & Policy*, vol. 2, e4, 2020.

[9] D. Tith et al., "Patient consent management by a purpose-based consent model for electronic health record based on blockchain technology," *Healthcare Informatics Research*, vol. 26, no. 4, pp. 265–273, 2020.

[10] T. van den Broek and A. F. van Veenstra, "Modes of governance in inter-organizational data collaborations," in *ECIS 2015 Completed Research Papers*, 2015, pp. 0–12. DOI: 10.18151/7217509.

[11] I. Jussen et al., "Issues in inter-organizational data sharing: Findings from practice and research challenges," *Data & Knowledge Engineering*, vol. 150, p. 102 280, 2024, ISSN: 0169-023X.

[12] J. S. Park, R. Sandhu, and G.-J. Ahn, "Role-based access control on the web," *ACM Transactions on Information and System Security (TISSEC)*, vol. 4, no. 1, pp. 37–71, 2001.

[13] R. Abdunabi, M. Al-Lail, I. Ray, and R. B. France, "Specification, validation, and enforcement of a generalized spatio-temporal role-based access control model," *IEEE Systems Journal*, vol. 7, no. 3, pp. 501–515, 2013. DOI: 10.1109/JSYST.2013.2242751.

[14] V. C. Hu et al., "Guide to attribute based access control (abac) definition and considerations (draft)," *NIST special publication*, vol. 800, no. 162, pp. 1–54, 2013.

[15] V. C. Hu, D. R. Kuhn, D. F. Ferraiolo, and J. Voas, "Attribute-based access control," *Computer*, vol. 48, no. 2, pp. 85–88, 2015.

[16] B. S. Radhika, N. V. N. Kumar, and R. K. Shyamasundar, "Towards unifying rbac with information flow control," in *Proceedings of the 26th ACM Symposium on Access Control Models and Technologies*, 2021, pp. 45–54. DOI: 10.1145/3450569.3463570.

[17] B. S. Radhika, N. V. N. Kumar, and R. K. Shyamasundar, "Samyukta: A unified access control model using roles, labels, and attributes," in *Information Systems Security*, Cham:

Springer Nature Switzerland, 2022, pp. 84–102, ISBN: 978-3-031-23690-7. DOI: 10.1007/978-3-031-23690-7_5.

[18] A. Karandikar, "What makes consent meaningful?" In *Companion Publication of the 16th ACM Web Science Conference*, 2024, pp. 42–46.

[19] M.-R. Ulbricht and F. Pallas, "Comafeds: Consent management for federated data sources," in *2016 IEEE International Conference on Cloud Engineering Workshop (IC2EW)*, 2016, pp. 106–111. DOI: 10.1109/IC2EW.2016.30.

[20] M. Casassa Mont, V. Sharma, and S. Pearson, "Encore: Dynamic consent, policy enforcement and accountable information sharing within and across organisations," HP Laboratories Technical Report, Tech. Rep., 2012.

[21] E. Olca and O. Can, "Dicon: A domain-independent consent management for personal data protection," *IEEE Access*, vol. 10, pp. 95 479–95 497, 2022. DOI: 10.1109/ACCESS.2022.3204970.

[22] J. Hyysalo, H. Hirvonsalo, J. Sauvola, and S. Tuoriniemi, "Consent management architecture for secure data transactions," Jul. 2016. DOI: 10.5220/0005941301250132.

[23] J. Asswad and J. Marx Gómez, "Data ownership: A survey," *Information*, vol. 12, no. 11, 2021, ISSN: 2078-2489. DOI: 10.3390/info12110465.

[24] D. Hart, "Ownership as an issue in data and information sharing: A philosophically based review," *Australasian Journal of Information Systems*, vol. 10, no. 1, Nov. 2002. DOI: 10.3127/ajis.v10i1.440.

[25] Ministry of Electronics and Information Technology, Government of India, *The digital personal data protection act, 2023*, https://www.meity.gov.in/static/uploads/2024/06/2bf1f0e9f04e6fb4f8fef35e82c42aa5.pdf, No. 22 of 2023. Gazette of India, August 11, 2023, 2023.

[26] A. He and R. Arcesati, "Data marketplaces and governance: Lessons from china," *Centre for International Governance Innovation (CIGI)*, 2023. [Online]. Available: https://www.cigionline.org/articles/data-marketplaces-and-governance-lessons-from-china/.

[27] Nordic Institute for Interoperability Solutions, *X-road® technology overview*, Accessed: 2024-12-17, 2024. [Online]. Available: https://x-road.global/x-road-technology-overview.

[28] A. Kalja, "The x-road project," *A project to modernize Estonia's national databases. Baltic IT&T review*, vol. 24, pp. 47–48, 2002.

[29] Web Science Lab, IIIT Bangalore, "Anumati consent management system," 2026. [Online]. Available: https://anumati.iiitb.ac.in/.

Collaborative Human-"AI ageNt Swarm" for Conducting Scientific Research

Wilbert Villalobos
Department of Computer Science and Technology
Kean University
Union, NJ
villalow@kean.edu

Yulia Kumar
Dept. of CS and Technology, Kean University
Department of ECE, Rutgers University
Union, NJ and Piscataway, NJ
ykumar@kean.edu

J. Jenny Li
Department of CS and Technology
Kean University
Union, NJ
juli@kean.edu

Dov Kruger
Department of Electrical and Computer Engineering
Rutgers University
Piscataway, NJ
Dov.Kruger@rutgers.edu

Jose Marchena
Department of CS and Technology
Kean University
Union, NJ
marchenj@kean.edu

***Abstract*—AI agents backed by Large Language Models (LLMs) can accelerate research tasks such as literature review, experiment prototyping, and technical writing; however, prompt-only workflows often provide limited provenance, weak reproducibility, and few safeguards against unsupported claims. This paper presents CHAINS, a Human-In-The-Loop (HITL) multi-agent research environment designed for accountable, inspectable research runs in Digital Society Systems (DSS) contexts. CHAINS decomposes a topic into structured stages—planning, paper retrieval and note normalization, gap formulation, micro-experiment execution with approval, drafting, critique, and deterministic verification. Each run produces a deterministic manifest that links stages to typed artifacts, including plans, notes, code, logs, plots, drafts, and verification reports, with hashes and per-step telemetry, enabling audit, replay, and cost/latency accounting. A verifier component performs non-generative checks over the draft and artifact bundle and emits a structured issue list to support remediation before release. We demonstrate the system through an end-to-end walkthrough and specify an evaluation protocol comparing CHAINS to an LLM-only baseline under matched budget constraints using workflow completion, artifact bundle integrity, grounding accuracy, HITL safety efficiency, and reproducibility.**

***Index Terms*—*Agentic AI*, *MCP*, *Collaborative Human-AI ageNt Swarm (CHAINS)*, *AGI*, *Digital Society Systems*, *Human-In-The-Loop*.**

I. Introduction

Digital-society systems (DSS) increasingly depend on rapid, evidence-based decision cycles: public-service agencies must draft policy briefs and service FAQs under tight timelines; municipalities must summarize citizen feedback and misinformation trends from participatory channels; and organizations must document decisions in a way that is auditable and reproducible. At the same time, modern scientific and software workflows are being supplemented by autonomous and semi-autonomous agentic systems that can search literature, draft analyses, and execute code. While such systems can expand what a single analyst or researcher can accomplish, they also introduce well-documented limitations—brittle reasoning, inconsistent grounding, and failure modes that produce plausible but incorrect outputs when not properly monitored [2]–[4]. For the Digital Society context, these failure modes are amplified by governance requirements: stakeholders need transparency, traceability, and safeguards that keep humans accountable for claims and actions [16], [17].

This paper presents CHAINS (Collaborative Human-AI ageNt Swarm), a Human-In-The-Loop (HITL) agentic laboratory designed for auditable, replayable knowledge work in digital-society settings. CHAINS is built around two engineering commitments. First, it is artifact-driven: each stage persists typed outputs—plans, structured literature notes, gap statements, experiment code/logs/plots when approved, drafts, and verification reports—and links them through an inspectable run ledger. Second, it enforces explicit HITL feasibility checkpoints that pause execution before code or tool actions, non-trivial costs, or high-impact claims. Given a topic, CHAINS executes a structured workflow—search and filtering, note normalization, gap formulation, optional micro-experiment execution, evidence-aligned drafting, critique, and verification—while recording a complete artifact bundle that can be inspected and re-run under controlled settings. Figures 1 and 2 overview the deployed system and dashboard implementation [5], [18].

Digital Society use cases. We target two representative scenarios that require accountability beyond prompt-only generation. The first is evidence bundles for public-service communication, such as drafting a policy brief or service FAQ with traceable sources and an auditable revision history. The second is citizen-participation and transparency reporting, such as summarizing community input, identifying recurring concerns, and producing a structured report with citations, metrics, and a reproducible audit trail. These scenarios motivate why agentic systems must expose intermediate artifacts, provide governance controls, and support repeatable evaluation under

budget and time constraints.

Threat model and safeguards. In addition to content-quality risks such as unsupported claims, digital-society deployments must mitigate prompt injection from retrieved content, unsafe tool execution, and inadvertent leakage of sensitive data. CHAINS addresses these risks via tool allowlists and step-level approvals, sandboxed execution for micro-experiments, structured note schemas to constrain evidence ingestion, and a deterministic verifier that performs non-generative checks over artifacts, including required-structure checks, citation/claim linkage rules, and run-ledger integrity checks before final reporting.

Reproducibility without overclaiming determinism. Rather than asserting fully deterministic reproduction of LLM outputs, CHAINS focuses on replayable audit. The run ledger records model identifiers, prompts/templates, tool calls, retrieved source identifiers and timestamps, environment metadata for code execution, and cryptographic hashes of artifacts. This enables reviewers and practitioners to inspect, attribute, and re-execute workflow steps under controlled conditions and to quantify costs and latency per stage.

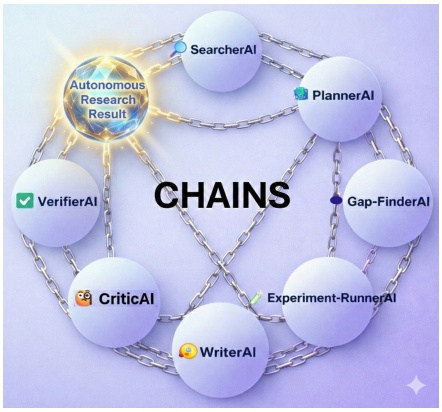


Fig. 1: CHAINS at-a-glance: artifact-driven agentic workflow.

Problem statement. Many current agentic workflows behave as "black boxes" that hide intermediate reasoning and evidence selection, blur provenance between retrieved sources and generated claims, and permit costly or unsafe tool actions without explicit governance. These issues hinder auditability and trust in digital-society deployments where decisions and communications must be accountable. We address the need for an artifact-first orchestration framework that couples HITL governance with structured verification and a replayable run ledger suitable for evaluation and oversight.

Research questions. We study three questions: (RQ1) Under matched budget constraints, how effectively can CHAINS translate a topic into actionable gaps and, when approved, feasible micro-experiments compared to baseline workflows? (RQ2) Does an artifact bundle and run ledger improve transparency and replayable audit compared to prompt-only and agentic baselines without systematic provenance? (RQ3) How do explicit HITL checkpoints and verifier-enforced rules reduce unsafe actions and unsupported claims?

Contributions. This paper makes five contributions: (1) a HITL agentic lab architecture tailored to auditable digital-society knowledge work; (2) an artifact schema and run-ledger design with hashes and metadata that supports inspection and replayable audit; (3) a verifier that performs deterministic, non-generative checks and emits structured issue lists for remediation; (4) an evaluation protocol with a prompt-only LLM baseline and an agentic-pipeline baseline using repeated trials and distributional reporting under equal budgets; and (5) two DSS case-study templates—public-service evidence bundles and citizen-participation transparency reports—with measurable outputs such as completion, artifact completeness, cost/latency, and unsupported-claim rate.

Finally, we position CHAINS relative to widely used agent frameworks and orchestration tooling, including graph-based agent workflows and agents SDK ecosystems, by clarifying what is enforced in CHAINS—typed artifacts, verifier-gated governance, and a replayable audit ledger—versus what is typically optional in general-purpose frameworks.

The remainder of the paper is organized as follows. Section II reviews AI-scientist systems, agent frameworks, evaluation protocols, and safeguards-first analyses. Section III describes the CHAINS architecture, agents, run ledger, operational deployment, threat model, and verification design. Section IV reports current prototype results under matched-budget framing. Section V presents a single-run walkthrough and formal evaluation metrics. Section VI discusses limitations and governance implications for digital-society deployments. Section VII concludes and outlines future extensions, including HPC-backed and quantum-inspired infrastructure.

II. Related Work

Agentic AI scientist systems increasingly couple literature retrieval, planning, code execution, experimentation, and drafting into integrated pipelines. Digital-society applications often demand accountable and inspectable outputs, such as reports for public services, citizen-facing summaries, and decision support. The central question is therefore not only whether an agent can generate a paper-shaped artifact, but whether the process is auditable, reproducible, and governed under explicit constraints. Accordingly, we organize related work by autonomy-first end-to-end research loops, evaluation benchmarks and protocol design, domain-specific co-scientists, and safeguards-first analyses.

Autonomy-first end-to-end research loops. The AI Scientist [12] and AI Scientist-v2 [11] exemplify systems optimized for autonomous iteration over ideas, code, experiments, and writing. These works are important reference points for maximizing capability and throughput; however, they are not primarily framed around accountability artifacts such as run manifests, hashed intermediate outputs, and explicit governance checkpoints that make a run inspectable and replayable under constrained conditions. In contrast, CHAINS treats artifact provenance, step-level telemetry, and user-controlled feasibility gates as first-class requirements for producing audit-ready outputs suitable for digital-society contexts.

Benchmarks and evaluation protocols for agentic research. A recurring limitation in the area is that success is often demonstrated via single exemplars rather than controlled protocols. AI-Researcher and Scientist-Bench [10] directly address this gap by formalizing tasks and evaluation criteria for whether an agentic system can implement or extend research methods. CHAINS adopts this benchmark-oriented perspective by defining workflow completion and artifact integrity as measurable endpoints, specifying a matched-budget LLM-only baseline, and tracking cost/latency per step to enable realistic comparisons when budgets and operational constraints matter.

Domain-specific co-scientists and high-stakes validation. Domain-centric co-scientists, such as biomedical discovery systems [14], emphasize specialized tooling, domain knowledge, and validation workflows. CHAINS is positioned differently: it is a domain-agnostic agentic laboratory substrate whose primary contribution is orchestration for accountable research runs—including typed artifacts, replayable manifests, and explicit approvals—rather than domain-specific discovery.

Surveys, commentary, and safeguards-first analyses. Surveys synthesize common components and bottlenecks across AI scientist systems [15], while commentary highlights the pace of progress and the risk of over-claiming acceleration without robust evidence [13]. Safeguards-first work argues that monitoring, constraints, and governance must be explicit because errors can propagate into downstream scientific and societal claims [16], [17]. CHAINS operationalizes these needs by producing a structured run record and by exposing failures as a deterministic issue list rather than as informal narrative. Table I summarizes these differences.

TABLE I: RELATED WORK COMPARISON.

Ref.	Focus	Key Differences vs. CHAINS
[12]	End-to-end automated research loop from idea to code, experiments, and paper.	Autonomy-first; accountability is not centered around replayable run manifests, per-step telemetry, or explicit governance gates for tool execution and release of claims.
[11]	Tree-search style agentic refinement over code and experiments.	Optimizes autonomous refinement; CHAINS emphasizes controlled orchestration, typed artifacts, and inspectable failure handling aligned with audit-ready reporting.
[10]	AI-Researcher and Scientist-Bench evaluation protocols for agentic research.	Benchmark-centric protocol design; CHAINS contributes an implementation substrate and aligns evaluation around repeated trials, matched-budget baselines, and artifact integrity.
[14]	AI co-scientist for biomedical discovery.	Domain-specific validation; CHAINS is domain-agnostic infrastructure emphasizing provenance, governance, and reproducible artifact bundles for broader digital-society use cases.
[15]	Survey of AI scientist systems and bottlenecks.	Taxonomy and roadmap; CHAINS instantiates a DSS-relevant accountability layer with manifests, hashes, telemetry, and deterministic verification.
[16]	Safeguards-first risk analysis for AI scientists.	Risk/governance analysis; CHAINS operationalizes safeguards via explicit HITL gates and deterministic verification over artifacts to reduce unsupported-claim risk.

III. METHODOLOGY

A. Run Model, Governance, and Operational Realism

A run is the unit of execution. Given a natural-language topic and a run mode, CHAINS advances through a fixed step sequence—plan, search, gaps, experiments, write, review, verify—while persisting intermediate artifacts and state transitions in a machine-readable run manifest. The current prototype is deployed as a web-accessible service, reflecting an operational setting rather than an offline script, and enabling realistic measurement of end-to-end latency, failure modes, and user interventions [5], [18].

Governance is enforced through explicit HITL gates at points where the system would otherwise execute code or tools, incur non-trivial budget consumption, or release externally facing claims. In the default configuration, CHAINS uses two mandatory checkpoints: Gate A requires approval before Experiment_RunnerAI may execute code or call external tools, and Gate B requires approval before publishing the final output bundle intended for stakeholders. These gates instantiate safeguards-first guidance by making high-risk actions deliberate and auditable rather than implicit [16], [17].

B. Agents, Responsibilities, and Typed Artifacts

CHAINS uses seven task-specialized agents. Each agent consumes and produces typed artifacts with an explicit schema, enabling provenance, replay, and deterministic checks. This role separation reduces cross-task contamination, such as drafting before evidence is stabilized, and supports controlled transitions with HITL gates. The agent roles are summarized in Table II.

TABLE II: CHAINS AGENTS.

Agent	Main Function
SearcherAI	Formulates literature queries, retrieves sources, and produces structured notes with problem, method, dataset, results, limitations, and identifiers/URLs.
PlannerAI	Produces a run plan with objectives, section outline, evidence acceptance criteria, and budget constraints for downstream steps.
Gap_FinderAI	Synthesizes literature notes into one to three actionable gaps and proposes feasible micro-experiments under the declared constraints.
Experiment_ RunnerAI	After Gate A approval, executes micro-experiments, emits logs/plots/CSVs, and records runtime environment details relevant for reproduction.
WriterAI	Drafts a manuscript or stakeholder document grounded in available artifacts only, linking claims to notes and experiment artifacts.
CriticAI	Performs an LLM-based critique focused on ambiguity, unsupported claims, and missing evidence links.
VerifierAI	Runs deterministic checks and produces a structured issue list that drives remediation and supports Gate B decisions.

C. Deterministic Verification and Failure Handling

To reduce reliance on self-consistency via another LLM, CHAINS includes a deterministic verification pass that runs

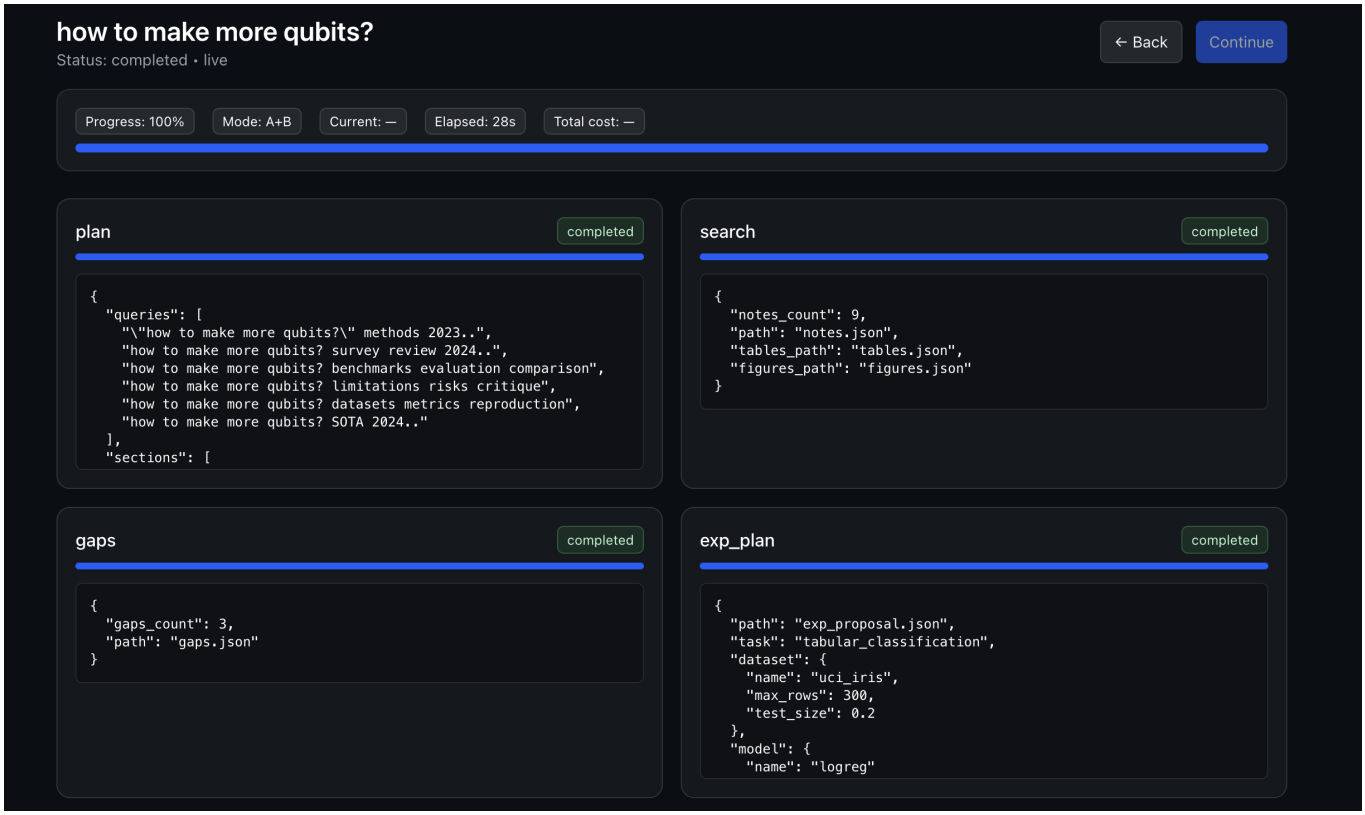


Fig. 2: CHAINS dashboard for run monitoring, artifact inspection, and HITL approvals [5].

over the artifact bundle and draft. The verifier produces a structured report, *verify.json*, containing pass/fail status, numeric diagnostics, and an actionable issue list grouped by severity. In the current prototype, checks include artifact integrity, traceability, citation consistency, and quality constraints such as required sections and configurable length thresholds. Failures are not treated as a single terminal state; instead, the verifier emits repair targets that can be routed back to WriterAI and CriticAI under governance, making remediation explicit and repeatable.

D. System Implementation, API Surface, and Reproducibility Hooks

CHAINS uses a FastAPI/Uvicorn backend and a Next.js/Tailwind frontend, integrating OpenAI agent tooling [6]. Event streaming ensures real-time updates rendered through the dashboard. The backend exposes endpoints for run creation, snapshot retrieval, and HITL approvals, enabling realistic interactive use, controlled evaluation runs, and reproducibility via standardized execution semantics. Table III lists representative endpoints.

Figure 3 shows the system architecture. A researcher interacts with the web UI, which communicates with the FastAPI backend. The OrchestratorAI coordinates CHAINS pipelines, calls MCP-style tools, stores artifacts in filesystem-based run directories, integrates with external AI/data services, and supports HITL approvals. The backend currently runs on a Microsoft Azure virtual machine and can optionally use Azure Blob Storage for cloud object storage, allowing the system to scale from local deployment to distributed artifact storage [7], [8]. Pydantic validation defines request and response models at the HTTP boundary, including run-creation requests, approval payloads, snapshots, step records, artifact records, progress summaries, and cost summaries [9].

TABLE III: SELECTED API ENDPOINTS.

Method, Endpoint	Description
POST /runs	Creates a run and advances until a pause for HITL approval or completion; returns a snapshot for UI rendering.
GET /runs/{topic}	Returns the latest snapshot and run-manifest summary for a topic.
GET /runs	Lists recent runs with status, progress, and telemetry summary.
POST /runs/{topic}/continue	Resumes a paused run to the next gate or terminal state.
POST /runs/{topic}/approve	Records approval or rejection for a waiting step and advances the run.

E. Run Manifest, Provenance Granularity, and Replayability

The run manifest is the single source of truth for provenance and replay. It records run metadata such as topic, mode, timestamps, and configuration; step state transitions; artifact ledger entries; telemetry summaries; and failure/remediation records. Each artifact is represented as an *ArtifactRecord* with path, kind, size, hash, and metadata. Steps are tracked

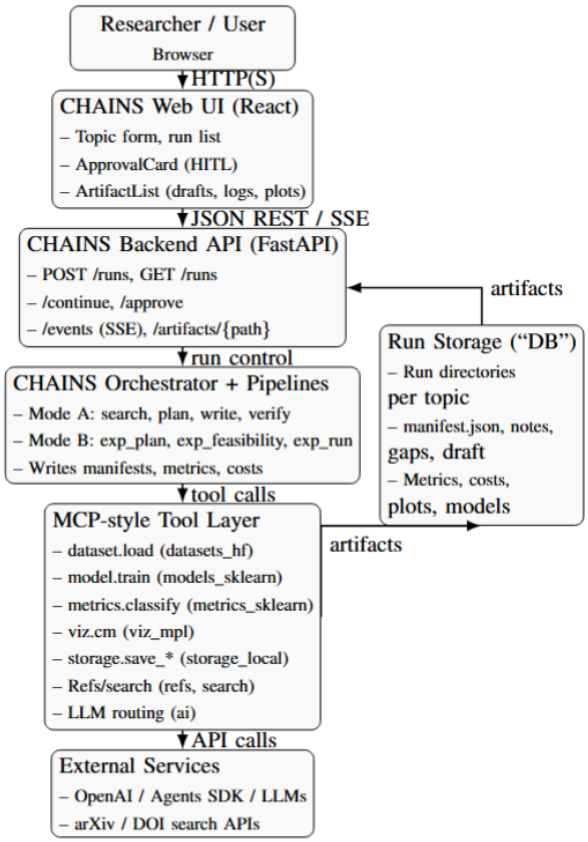


Fig. 3: High-level architecture of the CHAINS system.

with *StepRecord* entries, including status, timestamps, and error details. This structure supports two reproducibility goals: replay, which reconstructs a run state using the manifest plus the artifact directory; and audit, which inspects decision points, HITL approvals, evidence links, and deterministic verification outcomes.

F. Telemetry, Budgeting, and Matched-Baseline Support

To enable protocol-ready evaluation and fair comparisons, CHAINS logs per-step telemetry, including wall-clock duration, tool/model calls, and token/cost accounting. These signals are written as part of the run manifest and exported as CSV/JSON for analysis. Budgeting is enforced at the plan level and tracked at runtime. The same schema is used to define a matched-budget LLM-only baseline that receives an equivalent budget envelope but does not benefit from the multi-agent artifact pipeline. This directly supports controlled comparisons on completion rate, artifact completeness, and unsupported-claim incidence.

G. Inputs, Outputs, and Digital-Society Output Bundles

CHAINS accepts a minimal topic plus a mode, such as research memo, policy brief, or citizen FAQ, and produces a structured output bundle intended for stakeholders. The bundle is generated only after deterministic verification and Gate B approval. It includes a primary artifact, a transparency summary explaining sources and uncertainties, a reproducible artifact pack containing the run manifest, hashes, notes, gaps, logs, and telemetry, and a verification report with deterministic checks and issue lists. This packaging aligns agentic research outputs with evidence workflows that prioritize traceability and public accountability.

H. Episode Memory and Rubric-Based Patching

CHAINS currently focuses on accountable single-run execution. Ongoing work adds cross-episode improvement while preserving auditability: episodic memory storing outcomes and failures as structured JSON, semantic memory indexing prior artifacts for retrieval, and a rubric-based grader scoring each run against predefined criteria such as traceability, completeness, and constraint compliance. Rubric results and hard metrics, including accuracy, cost, latency, and violations, produce a compact patch artifact that updates prompts, routing, retrieval filters, and tool-usage constraints for subsequent runs, keeping optimization aligned with safeguards-first priorities [16], [17].

IV. Current Results

CHAINS has been implemented as a working, web-accessible prototype and deployed with a live dashboard for run monitoring and HITL control [5]. In the current deployment, each run executes a fixed, auditable step sequence—planning, literature retrieval and normalization, gap identification, optional micro-experiment execution after approval, drafting, critique, and deterministic verification—while persisting intermediate outputs as typed artifacts and recording all state transitions in a run manifest. The dashboard exposes step status, artifacts, and gate decisions so that cost- and risk-bearing actions are explicit, reviewable, and reversible before publication.

A. System-Level Stability and Artifact Production

Across multiple modest topics representative of early-stage inquiry, CHAINS reliably completes end-to-end orchestration and produces a consistent artifact bundle: a machine-readable plan, structured literature notes, actionable gaps, a draft scaffold, deterministic verification output, and telemetry/budget traces. The key result at this stage is not claim novelty but accountable execution: intermediate decisions are preserved, artifacts are hash-addressable, and failures are localized to a specific step with a structured issue list rather than hidden inside a monolithic prompt-response interaction.

B. Comparison to an LLM-Only Baseline

In preliminary side-by-side trials under comparable budget envelopes, CHAINS more consistently completed the intended workflow and yielded more structured, inspectable outputs than an LLM-only baseline. The baseline frequently produced paper-shaped text with unclear provenance, missing intermediate artifacts, and a higher incidence of unsupported assertions. By contrast, CHAINS enforces explicit checkpoints and preserves an artifact ledger that supports inspection and replay, allowing users to trace claims to sources or experimental outputs and identify which step introduced an error.

HITL gates, especially approvals before code/tool execution and before final release, reduce risky or wasteful actions and provide a governance interface aligned with safeguards-first guidance [16], [17].

C. Limitations of Current Evidence

Current evidence is primarily system-level: stability of orchestration, completeness of artifact bundles, and verifiable run accounting. These results do not yet constitute large-scale benchmarking or statistically supported claims about general research acceleration. Controlled evaluation over multiple topics, repeated trials, and distributional reporting of cost/latency and failure modes remains ongoing work and is required for stronger quantitative conclusions.

V. Case Study: Walkthrough of a Single Run

To make CHAINS concrete, we report an end-to-end run on the topic "best ways to detect sounds." The run produced an inspectable artifact bundle—plan, literature notes, gaps, draft scaffold, deterministic verification report, and telemetry—and recorded hashes and sizes for each artifact in the run manifest, enabling replay, audit, and failure localization. The run executed the standard pipeline and completed in 4.834 seconds wall-clock time end-to-end, with per-step durations recorded.

Planning. CHAINS converts the topic into a structured plan containing explicit search queries, writing objectives, evidence acceptance criteria, and a section-by-section outline with concrete TODO items. The plan also specifies run constraints such as budget and optional experiment policy, ensuring downstream steps produce consistent outputs even when content quality varies.

Retrieval and note normalization. The search step produces structured notes for each selected source, including title, authors, date, identifier/URL, short summary, and placeholders for datasets, results, and limitations. This validates the note schema and persistence layer and makes retrieval outputs directly consumable by downstream agents.

Gap synthesis. The pipeline emits a valid *gaps.json* artifact consisting of one to three actionable gaps and feasible experiment proposals. Gaps are stored as structured objects with fields designed for execution planning and later evaluation, including feasibility constraints, required data, and expected measurable outcomes.

Drafting, critique, and deterministic verification. The drafting stage produces substantive body text and a structured skeleton, allowing the UI and downstream verifiers to diagnose failures precisely. CHAINS then performs a deterministic verification pass and emits a structured report containing status, numeric diagnostics, and an actionable issue list. Quality control is therefore not delegated solely to an LLM critic; the system includes a rule-based checker that can gate release and drive targeted remediation.

Telemetry and accounting. The run logs capture end-to-end and per-step wall-clock time and token/call accounting per model and step. Even when content is incomplete, the system still produces a complete cost ledger and timing trace, enabling realistic comparisons between autonomy-first workflows and governed, HITL workflows under budget constraints.

A. Evaluation Protocol and Metrics

To enable stronger statistical claims, we evaluate each topic across repeated runs under identical constraints and compare CHAINS against an LLM-only baseline under matched budget envelopes. Outcomes are derived from the run manifest, artifact schemas, and verifier reports. We summarize the evaluation with a composite score that aggregates five primary metrics:

$$\text{Score} = w_1\text{WCR} + w_2\text{ABI} + w_3\text{GAS} + w_4\text{HSE} + w_5\text{RI}, \quad (1)$$

where the weights satisfy $w_i \geq 0$ and $\sum_{i=1}^{5} w_i = 1$, and each metric is normalized to $[0, 1]$.

Workflow Completion Rate (WCR) is the fraction of initiated runs that reach deterministic verification without terminal failure. Artifact Bundle Integrity (ABI) is the fraction of runs whose required artifacts—planning, literature, draft, telemetry, and verification—are present and schema-valid. Grounding Accuracy Score (GAS) is the fraction of draft claims that resolve to evidence anchors in notes and/or experiment artifacts. HITL Safety Efficiency (HSE) is a normalized count of high-risk actions intercepted or corrected by HITL gates, such as blocked tool execution or revised claim release. Reproducibility Index (RI) is the fraction of runs for which the run state can be reconstructed using only the manifest, hashes, and artifact directory. In experiments, we report both individual metrics and the composite score in Eq. (1) to avoid masking specific failure modes while still providing an overall measure of accountable workflow performance under matched budgets.

TABLE IV: SYSTEM PERFORMANCE COMPARISON.

Metric	LLM-Only	CHAINS
WCR	0.40	0.90
ABI	0.00	1.00
GAS	N/A	Higher; anchor-checked
HSE	0.00	2+ Gate A/B interventions
RI	0.00	1.00; manifest + hashes

VI. Discussion

While CHAINS demonstrates a working, accountable agentic research workflow, several limitations and governance implications are important for interpreting the current results and positioning the system for DSS use cases. First, CHAINS does not eliminate upstream evidence risk: retrieval can select incomplete, outdated, or low-quality sources, and structured notes may still omit critical limitations unless explicitly required by the rubric. For DSS contexts, this motivates stricter evidence policies such as source-quality tiers, date constraints, and mandatory limitation fields, as well as stronger verifier rules that reject claims lacking evidence anchors.

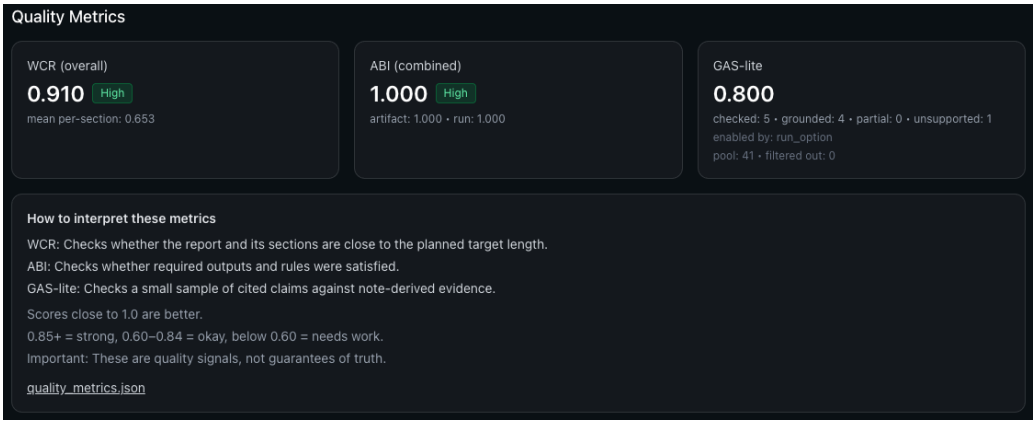
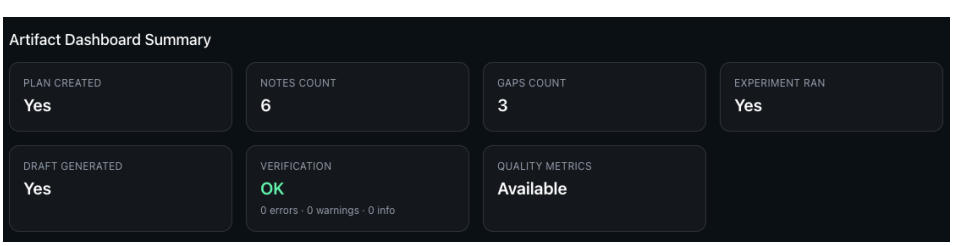
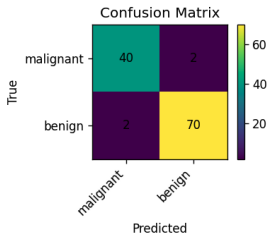


Fig. 4: CHAINS quality metrics panel showing interpretable artifact-level quality signals.



(a) Post-run artifact inspection view.



(b) Micro-experiment result.

Fig. 5: Interpretable CHAINS results: high-level artifact summary and experiment-level diagnostics.

Second, deterministic verification currently focuses on structural integrity and traceability checks; it does not guarantee semantic correctness of claims, and it cannot fully prevent subtle misinterpretations of cited work. CHAINS should therefore be treated as a decision-support and reporting substrate, not an autonomous authority. Third, HITL gates reduce risk but introduce human workload. Poor gate design can either overburden users or provide a false sense of safety. In practice, governance should be configurable by deployment setting, with stricter gating for public-facing outputs and lighter gating for internal exploration. Finally, budget and telemetry logging improve accountability but create a new responsibility: teams must define acceptable cost/risk envelopes and retention policies for artifact bundles that may include sensitive content.

VII. Conclusion and Future Work

This paper introduced CHAINS, an artifact-driven, human-governed agentic laboratory for research workflows in which accountability is a first-class requirement. CHAINS orchestrates task-specialized agents across planning, literature retrieval and normalization, gap identification, optional micro-experiment execution, drafting, critique, and deterministic verification, while persisting every intermediate output as typed artifacts indexed by a run manifest. Unlike prompt-only workflows that are difficult to audit, CHAINS makes provenance and reproducibility operational: runs are replayable from a manifest and hash-checked artifact ledger, failures are localized to explicit steps, and governance is enforced via HITL feasibility gates before tool/code execution and before final release. This safeguards-first design supports ICDS-style digital-society evidence workflows that must produce transparent outputs under real-world constraints on cost, latency, and risk.

A. Future Work

Our next steps focus on strengthening CHAINS along the dimensions that matter for trustworthy deployment and evaluation in digital society settings. First, we will expand controlled benchmarking under realistic constraints with repeated trials across a topic suite, matched-budget baselines, and distributional reporting of completion, cost/latency, and failure modes. Second, we will extend deterministic verification from structural checks to richer accountability checks, including stricter claim-to-evidence linking, citation/identifier validation, and policy-driven constraints on tool usage and external calls. Third, we will broaden server-side integrations, datasets, domain tools, and specialized or fine-tuned models while preserving governance semantics. Fourth, we will implement cross-episode learning with audit-preserving updates, including episodic memory, semantic memory over prior artifacts, and rubric-based patch artifacts. Finally, we will strengthen stakeholder-facing packaging for DSS use cases through standardized release bundles containing a primary deliverable, transparency summary, reproducible artifact pack, and verifier report.

B. HPC and Quantum-Enabled Extensions for CHAINS

Beyond software-only improvements, we propose two infrastructure upgrades—HPC-backed execution and quantum-enabled retrieval/optimization—to expand the scope of micro-experiments, improve evaluation rigor, and enable new classes of agentic workflows while preserving governance and auditability.

HPC-backed micro-experiments and scalable benchmarking. A key limitation of current agentic lab systems is that experiments are typically constrained to local resources, which biases evaluation toward small tasks and discourages repeated trials. We will integrate CHAINS with an HPC environment, such as a Slurm-based cluster, by introducing an execution adapter that converts approved experiment plans into batch jobs. Under this design, Experiment_RunnerAI does not directly execute heavy workloads; instead, after HITL approval it submits a job bundle consisting of a container or environment specification, a deterministic entrypoint script, and an input artifact manifest. The scheduler returns a job ID that is tracked in the run manifest, and all outputs—logs, metrics, plots—are collected into the artifact directory upon completion. This enables larger topic suites, more repetitions per topic, environment capture with fixed seeds, strict queue policies, and explicit failure localization for scheduler errors or resource exhaustion.

Quantum-enabled retrieval and optimization under budget constraints. We further propose exploring quantum and quantum-inspired components as optional, governance-controlled modules within CHAINS. The goal is not to claim universal speedups, but to provide a testable pathway for hybrid classical–quantum workflows where quantum resources are treated as scarce, auditable tools. Near-term integration targets include quantum-assisted retrieval/reranking, where candidate documents are mapped into an embedding space and quantum or quantum-inspired similarity scoring reranks top-k candidates, and quantum-inspired search for agent routing and experiment selection, where tool selection and experiment scheduling are modeled as constrained discrete optimization problems. Solver outputs will be treated as recommendations subject to HITL approval, and the manifest will record the objective, constraints, solver configuration, backend mode, shot count when applicable, and selected action set.

Adding HPC and quantum modules enables additional DSS-relevant metrics beyond Eq. (1). We summarize these dimensions with an extended composite score:

$$\mathrm{Score}_{\mathrm{HPCQ}} = \alpha\, w^{\top} m + (1-\alpha)\, v^{\top} r, \tag{2}$$

where $\alpha \in [0,1]$, $\sum_i w_i = 1$, $\sum_j v_j = 1$, and all weights are nonnegative. The vectors $\mathbf{m} = \big[\mathrm{WCR},\ \mathrm{ABI},\ \mathrm{GAS},\ \mathrm{HSE},\ \mathrm{RI}\big]^{\top}$ and $\mathbf{r} = \big[\mathrm{CE},\ \mathrm{EC},\ \mathrm{HMR},\ \mathrm{GL}\big]^{\top}$ capture core accountability metrics and resource/governance metrics, respectively. CE is Compute Efficiency, EC is an Energy/Carbon proxy, HMR is Hardware-Mode Robustness across local/HPC/quantum modes, and GL is Governance Load, defined as inverse-normalized human time and approvals per run. Overall, CHAINS is an architectural step toward reducing fragmentation in modern research practice by unifying orchestration, provenance, telemetry, verification, and governance in one operational pipeline.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation CISE Graduate Fellowships under Grant No. 2313998. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] T. S. Ashton, *The Industrial Revolution 1760–1830*. Oxford, U.K.: Oxford University Press, 1997.

[2] H. L. Dreyfus, *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge, MA, USA: MIT Press, 1992.

[3] T. Hagendorff and K. Wezel, "15 challenges for AI: Or what AI (currently) can't do," *AI & Society*, vol. 35, no. 2, pp. 355–365, 2020.

[4] K. Crawford and R. Calo, "There is a blind spot in AI research," *Nature*, vol. 538, no. 7625, pp. 311–313, 2016, doi: 10.1038/538311a.

[5] Agents Laboratory, "Agents Laboratory—A+B demo," 2025. [Online]. Available: https://app.agents-lab-experiments-project.com/. Accessed: Nov. 27, 2025.

[6] OpenAI, "Agents SDK," OpenAI Platform Documentation, 2025. [Online]. Available: https://platform.openai.com/docs/guides/agents-sdk. Accessed: Nov. 27, 2025.

[7] Microsoft Azure, "Azure Virtual Machines," 2025. [Online]. Available: https://azure.microsoft.com/services/virtual-machines/. Accessed: Nov. 29, 2025.

[8] Microsoft Azure, "Azure Blob Storage," 2025. [Online]. Available: https://azure.microsoft.com/services/storage/blobs/. Accessed: Nov. 29, 2025.

[9] S. Colvin et al., "Pydantic: Data validation and settings management using Python type hints," 2025. [Online]. Available: https://github.com/pydantic/pydantic. Accessed: Nov. 27, 2025.

[10] J. Tang, L. Xia, Z. Li, and C. Huang, "AI-Researcher: Autonomous scientific innovation," *arXiv preprint arXiv:2505.18705*, 2025.

[11] Y. Yamada, R. T. Lange, C. Lu, S. Hu, C. Lu, J. Foerster, J. Clune, and D. Ha, "The AI Scientist-v2: Workshop-level automated scientific discovery via agentic tree search," *arXiv preprint arXiv:2504.08066*, 2025.

[12] C. Lu, C. Lu, R. T. Lange, J. Foerster, J. Clune, and D. Ha, "The AI Scientist: Towards fully automated open-ended scientific discovery," *arXiv preprint arXiv:2408.06292*, 2024.

[13] D. Castelvecchi, "Researchers built an 'AI Scientist'—what can it do?" *Nature*, vol. 633, no. 8029, pp. 266–266, 2024.

[14] J. Gottweis et al., "Towards an AI co-scientist," *arXiv preprint arXiv:2502.18864*, 2025.

[15] Q. Xie et al., "How far are AI scientists from changing the world?" *arXiv preprint arXiv:2507.23276*, 2025.

[16] X. Tang et al., "Risks of AI scientists: Prioritizing safeguarding over autonomy," *Nature Communications*, vol. 16, no. 1, Art. no. 8317, 2025.

[17] X. Tang et al., "Prioritizing safeguarding over autonomy: Risks of LLM agents for science," in *Proc. ICLR Workshop on Large Language Model (LLM) Agents*, 2024.

[18] W. Villalobos, "CHAINS: Collaborative Hybrid AI-Researcher Networked System," GitHub repository, 2025. [Online]. Available: https://github.com/WilbertFV/CHAINS. Accessed: Nov. 29, 2025.

QoLI$_{v2}$: A Data-Driven Model for Comprehensive Quality of Life Assessment

Ilie Cristian Dorobăț
Faculty of Computer Science
National University of Science and Technology POLITEHNICA Bucharest
Bucharest, Romania
email: ilie.dorobat@upb.ro

***Abstract*—The assessment of Quality of Life (QoL) is a highly relevant and increasingly prominent topic, with numerous general or domain-specific metrics, either health-related or not, being developed to capture various levels of specificity in evaluating QoL. Following an in-depth review of some of the most widely recognized metrics that embrace the multidimensional nature of QoL, we identified several limitations that hinder the evaluation process. These encompass the limited number of incorporated indicators, the methods used for their aggregation, and the lack of a standardized framework to automate the assessment process. Consequently, we aimed to improve the Quality of Life Index (QoLI), initially introduced in 2019, by significantly expanding the number of indicators (from 32 to 76 in the updated QoLI$_{v2}$), revising the aggregation formula for its nine dimensions (eight objective dimensions and one subjective dimension) and their underlying indicators to produce more meaningful results. To evaluate the construct validity of QoLI$_{v2}$, Spearman correlation analyses were performed against two established benchmark indicators, Gross Domestic Product per Capita (GDP per Capita) and the Human Development Index (HDI). Strong monotonic relationships were observed with GDP per Capita (ρ = 0.847) and HDI (ρ = 0.804), confirming that QoLI$_{v2}$ is consistent with established development metrics while supporting its credibility as a multidimensional measure of quality of life.**

Keywords-Eurostat; Qualify of Life; Welfare; Well-being.

I. INTRODUCTION

The assessment of Quality of Life (QoL) has gained increasing visibility in recent years, with numerous studies proposing a wide range of metrics, both general and domain-specific, health-related or otherwise, for evaluating the population's QoL [1][2]. Although the term QoL has not yet received a universally accepted definition, nor is there a unified methodology for its assessment [3][4][5], it is generally regarded as a complex construct encompassing both objective (descriptive) and subjective (evaluative) factors that significantly influence individual and collective well-being.

Gross Domestic Product (GDP) has long been employed as the principal indicator for assessing population welfare [6][7][8], due to its ease of calculation and application as a macroeconomic measure. However, even its derivative, GDP per Capita, fails to provide a comprehensive view of societal well-being, as GDP primarily reflects the community's capacity to produce goods and services. Since the concept of QoL transcends the financial boundaries of GDP, researchers have increasingly turned their attention toward proposing multidimensional approaches for assessing QoL [9].

A wide range of studies from diverse branches of medicine illustrate this trend, ranging from the analysis of factors contributing to dental diseases and their associated consequences (e.g., pain, discomfort, impaired physical function) [10][11][12], to the examination of cancer-related variables and their influence on the well-being of patients [13], and to investigations into determinants of successful aging and strategies for enhancing the QoL in older adults [14][15]. These represent just a few examples from the medical field where researchers have developed targeted QoL metrics.

However, the development of such indicators extends well beyond the healthcare field. For instance, certain studies address the economic, social, cultural, and environmental impacts of tourism on residents' well-being [16][17], while others highlight the effects of urban development on quality of life [18][19]. In addition to these aggregated indicators that explore the well-being of the population within a specific domain, there is also a series of benchmark indicators that capture the complex and multivariate nature of life.

The Human Development Index (HDI) is an indicator developed by the *United Nations Development Programme* in 1990 and revised in 2010, which integrates three essential dimensions of human development: health, education, and a decent standard of living, aggregated using the geometric mean [20][21]. *The World Health Organization Quality of Life (WHOQOL)* is an initiative of the World Health Organization that assesses *"individuals' perceptions of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns"* [22], consisting of a set of 100 questions, organized into six distinct dimensions. *The Better Life Index (BLI)* is an indicator introduced in 2011 by the Organization for Economic Co-operation and Development (OECD) that encompasses measuring the material conditions of citizens, education level, health status, degree of security, social engagement, and other significant aspects influencing the population's QoL.

Unfortunately, although these indicators are widely used in assessing the population's quality of life, they present a series of limitations that constrain the scope of analyses and hinder the automation of calculations. The HDI integrates only three core dimensions, which may result in a partial representation of the multifaceted nature of well-being. The WHOQOL instrument relies on self-reported survey data, capturing individuals' subjective perceptions of well-being, which introduces methodological challenges related to response bias, cultural differences in self-evaluation, and cross-national comparability of aggregated results [22]. Regarding the BLI, OECD does not provide an official public calculation formula, and the published values refer exclusively to the organization's member states, thus limiting the global applicability of the indicator.

This paper aims to address the limitations of the aforementioned indicators by exploring a novel perspective on the methodology for assessing QoL, proposing and discussing in detail, in Section 2, a revised version of the Quality of Life Index (QoLI) formula [23], hereafter referred to as $QoLI_{v2}$. The updated formula addresses limitations identified in the initial version by revising the aggregation methodology and expanding the indicator set, while maintaining the same nine-dimensional structure of the index (eight objective and one subjective dimensions). Section 3 introduces the data sources, followed by an analysis of the correlation between the values produced by $QoLI_{v2}$ and two of the benchmark indicators: GDP and the HDI. Section 4 is devoted to outlining the study's conclusions.

II. $QoLI_{v2}$ METHODOLOGY

QoLI is a metric proposed in 2019 [23] that enables the measurement of population quality of life based on a broad range of factors. Unlike other well-known metrics such as HDI, which operationalizes three core dimensions using four specific indicators (hereafter referred to as atomic indicators), WHOQOL, which is derived from the application of a questionnaire, or BLI, which comprises 24 atomic indicators, the original version of QoLI included 32 atomic indicators selected from the set of indicators proposed by Eurostat [24] for assessing QoL.

$QoLI_{v2}$ expands the scope of the original QoLI to 76 atomic indicators by integrating measures of objective living conditions and subjective well-being evaluations, while introducing substantial revisions to the aggregation methodology. Unlike the original formula, which used the geometric mean to aggregate its dimensions, $QoLI_{v2}$ employs the logarithmic function for this purpose. Additionally, the new formula corrects errors and omissions identified in the initial version by updating, replacing, or adding new atomic indicators.

A. *The Structure of $QoLI_{v2}$*

The structure of $QoLI_{v2}$ is presented in Formula (1), where D represents the set of dimensions comprising $QoLI_{v2}$.

$$D = \{MLC, PMA, Education, Health, LSI, Safety, Governance, Environment, Overall\ Exp\} \quad (1)$$

where:
MLC = Material and Living Conditions dimension;
PMA = Productive or Main Activity dimension;
Education = Education dimension;
Health = Health dimension;
LSI = Leisure and Social Interactions dimension;
Safety = Economic and Physical Safety dimension;
Governance = Governance and Basic Rights dimension;
Environment = Natural and Living Environment dimension;
Overall Exp = Overall Experience of Life dimension;

Table 1 shows the indicators associated with the index's nine dimensions, where indicators listed in black have a positive connotation, while those shown in red have a negative connotation. To harmonize indicators that have a negative connotation, the *rev(x)* formula presented in Formula (6) is applied to them.

TABLE I. THE COMPOSITION OF THE 9 DIMENSIONS

Dimension Name	Indicator Name
Material and Living Conditions	Deprivation Ratio Dwelling Conditions Ratio Ends Meet Ratio Financial Satisfaction Ratio GDP per Capita Power High Income Ratio Income Quintile Ratio Lack of Baths Ratio Low Work Intensity Ratio Median Income Power Over Occupied Ratio Poverty Risk Ratio Under Occupied Ratio
Productive or Main Activity	Employment Ratio Inactive People Ratio Involuntary Part-Time Ratio Job Satisfaction Ratio Long Term Unemployed Ratio Low Wage Ratio Low Work Intensity Ratio Personal Time Left Researchers Ratio Temporary Employees Ratio Unemployed Ratio Working Flexibility Ratio Working Nights Ratio
Education	Digital Skills Ratio Dropout Ratio Early Education Ratio Education Ratio Inactive Young People Ratio No Knowledge of Any Foreign Language Ratio Pupils-to-Teachers Ratio Training Ratio $_{4w}$ Training Ratio $_{1y}$
Health	Body Mass Index Depressive Ratio

Dimension Name	Indicator Name
	Health Personnel Ratio Healthy Life Years Healthy People Ratio Hospital Beds Life Expectancy at Birth Long Term Medical Issues Ratio Non-Alcoholic Ratio Non-Fruits & Vegetables Ratio Physical Activities Ratio Smokers Ratio Unmet Dental Needs Ratio Unmet Medical Needs Ratio Work Accidents Ratio
Leisure and Social Interactions	Asking Ratio Discussion Ratio *Getting Together Ratio *Frequency Contact Ratio *Non-Participation in Events Ratio *Participation in Events Ratio Recreational Areas Satisfaction Ratio Relationships Satisfaction Ratio Time Satisfaction Ratio
Economic and Physical Safety	Crime Ratio Non-Payment Ratio Offences Ratio Pension Power Social Protection Power Unexpected Financial Expenses Ratio
Governance and Basic Rights	Citizenship Rate Gender Employment Gap Gender Pay Gap *Population Trust *Voter Turnout
Natural and Living Environment	*Air Pollution Ratio Noise Pollution Ratio Pollution Ratio Water Supply Ratio
Overall Experience of Life	Happiness Ratio Life Satisfaction Ratio

*Indicators marked with an asterisk are calculated as the average of the sub-indicators that compose them

B. *What's New in QoLI$_{v2}$ in Terms of The Structure of The Indicators?*

One of the most significant developments in QoLI$_{v2}$ concerns the refinement of the indicators used to calculate each individual dimension. A detailed account of these updates is provided below.

1) *Material Living Conditions (MLC)*

A new indicator, *GDP per Capita Power*, has been introduced in the calculation of this dimension to measure GDP per Capita in terms that are comparable across countries. By expressing GDP per Capita in terms of Purchasing Power Ptandards (PPS), a more balanced comparison can be made between countries with varying stages of economic development.

2) *Productive or Main Activity (PMA)*

The *Overqualified Ratio* indicator was removed from the aggregated calculation formula due to the lack of data provided by Eurostat. However, this exclusion was offset by the inclusion of two new indicators: the *Working Flexibility Rate* and the *Low Work Intensity Rate*. The former captures the proportion of individuals benefiting from a flexible work schedule, while the latter measures the share of people living in households with very low work intensity. It is important to note that although the latter indicator is already part of the aggregated formula for the *Material and Living Conditions* dimension, it has also been included in *Productive or Main Activity*, as Eurostat tracks the evolution of this indicator within both dimensions.

3) *Education*

The age range for calculating the *Inactive Young People Ratio* has been extended from individuals aged 18–24 to those aged 15–29 in order to better capture contemporary patterns of education-to-work transitions, align the indicator with international statistical standards, and provide a more comprehensive measure of youth inactivity across both early school-leaving and delayed labour market entry stages. Regarding the *Training Ratio*, in addition to the participation rate in education and training over the past four weeks, the participation rate over the past twelve months has also been introduced. Since the *Pupils to Teachers Ratio* reflects the number of primary and secondary students per teacher (the higher the number of students per teacher, the less individual time the teacher can allocate to each student), QoLI$_{v2}$ addresses the methodological oversight in the initial version by applying the reversed value formula.

4) *Health*

This dimension has been enhanced by introducing a new indicator: the *Depression Rate* and by replacing consumption-based indicators with 'non' rates to better capture exposure to health risk and allow a clearer identification of populations not meeting minimum healthy lifestyle standards. *Alcoholic Rate*, measuring the proportion of daily alcohol consumers, has been replaced by the *Non-Alcoholic Rate*, which reflects the proportion of individuals who have not reported any heavy episodic drinking in the past 12 months. *Fruits and Vegetables Consumers Rate*, measuring the proportion of the population that consumes fruits and vegetables daily, has been replaced by the *Non-Fruits and Vegetables Consumers Rate*, which captures the share of the population that does not consume any fruits or vegetables on a daily basis.

In the initial version of QoLI, the *Body Mass Index* was calculated by applying the reversed formula to the share of overweight and obese individuals, thereby overlooking underweight and pre-obese categories. In QoLI$_{v2}$, the indicator has been revised to measure the share of the population with a normal body weight. The *Health Personnel* indicator now includes midwives, and in order to align the scale of this indicator and that of *Hospital Beds* with others expressed as proportions or calendar years, their values are now reported per one million inhabitants instead of per 100,000.

5) *Leisure and Social Interactions (LSI)*

The indicators reflecting *the share of the population that regularly meets with family and friends* have been consolidated into a single measure, calculated using the geometric mean of the two original indicators. Similarly, *the participation rates in social activities, informal volunteering, and formal volunteering* have been merged into a single indicator, also calculated using the geometric mean. In

addition to these changes, *Frequency Contact Rate* has been introduced, capturing the share of the population that maintains regular contact with family or friends. The dimension was also enriched with the indicator *Satisfaction with Recreational and Green Areas*.

6) Economic and Physical Safety (Safety)

The *Offences* indicator has been enriched with statistics related to acts against computer systems, bribery, corruption, fraud, money laundering, organized criminal groups, and sexual exploitation.

7) Governance and Basic Rights (Governance)

The calculation of the *Voter Turnout* indicator has been enhanced by including not only parliamentary election results but also those of European Parliament and presidential elections. Additionally, for the *Gender Employment Gap* and *Gender Pay Gap* indicators, the calculation formulas have been adjusted to ensure that the presence of negative values (e.g., a higher proportion of men employed compared to women) does not result in an aggregated dimension value that is negative and lacks meaningful interpretation. This issue remained unnoticed in the initial version of QoLI, as both indicators consistently recorded values with the same sign, and their product therefore always resulted in a positive number.

8) Natural and Living Environment (Environment)

To provide a more comprehensive perspective on air quality, in addition to the population's exposure to particulate matter smaller than 2.5 µg/m³ and 10 µg/m³, the *Air Pollution* indicator has been enriched with data on exposure to acidifying gas emissions (NH3, NOX) and ozone precursors (CH4, CO, NMVOC, NOX).

9) Overall Experience of Life (Overall Exp)

This dimension has not undergone any changes.

C. $QoLI_{v2}$ Calculation Method

The approach used to calculate the $QoLI_{v2}$ value and its underlying dimensions has changed compared to the original formula, by applying the logarithmic function to the product of the aggregated parameters instead of using the geometric mean. This change was motivated by the asymmetric nature of these parameters [25], and by the fact that although the geometric mean also prevents a low value in one indicator from being offset by a high value in another [23], it provides results that are less interpretable than those produced by the logarithmic function [26].

We denote by *D* the set of all 9 dimensions that make up $QoLI_{v2}$, *d* a specific dimension from the set *D*, *e* an atomic indicator that is part of dimension *d*, and *e'* a sub-indicator of atomic indicator *e*. As shown in Formula (2), the calculation of $QoLI_{v2}$ involves applying the logarithmic function to the product of the values of the 9 constituent dimensions, while Formula (3) illustrates that each dimension is calculated using the same procedure, applied to the transformed values of its specific indicators.

$$QoLI_{v2} = \ln\left(\prod_{i=1}^{n} d_i\right) \ \forall\, d \in D,\ n = |D| \tag{2}$$

$$Dimension_{v2} = \ln\left(\prod_{i=1}^{n} t(e_i)\right) \ \forall\, e \in d,\ d \in D,\ n = |d| \tag{3}$$

To better clarify what is meant by *the transformed value of the indicators*, Formula (4) describes the method used to determine this value. Thus, three specific cases can be distinguished: i) the determination of values for composite atomic indicators; ii) the calculation of values for atomic indicators with a negative connotation; iii) the computation of values for regular atomic indicators.

$$t(x) = \begin{cases} avg(x), & x \text{ is a composite value} \\ rev(x), & x \text{ is a value with a negative connotation} \\ x, & \text{for other cases} \end{cases} \tag{4}$$

The first case refers to those atomic indicators that are composed of other indicators describing a related state, measured using the same unit. For these, the geometric mean described in Formula (5) is applied. For example, to determine the value of the *Voter Turnout* indicator, the geometric mean is applied to the indicators representing participation in European Parliament elections, parliamentary elections, and presidential elections.

$$avg = \sqrt[n]{\prod_{i=1}^{n} e'_n} \ \forall\, e' \in e,\ e \in d,\ d \in D,\ n = |e'| \tag{5}$$

Regarding the second case, a negatively connotated atomic indicator is one that expresses an adverse condition such as *the share of the population experiencing depressive symptoms*, *the proportion of inactive young people*, or *the percentage of the population that smokes*. Therefore, to avoid distorting the result by multiplying indicators that express positive states with those that reflect negative ones, Formula (6) allows for the transformation of the latter into positively connotated indicators. Thus, *the share of the population that experienced depressive symptoms* will be transformed into *the share that did not experience such episodes*; *the proportion of inactive young people* will be converted into *the share of young people who are not inactive*; *the percentage of smokers* will be replaced by *the percentage of non-smokers*, and so on.

$$rev(x) = 100 - x \tag{6}$$

Finally, if an atomic indicator is neither composite nor negatively connotated, it can be used as is, without any need for preprocessing.

III. Application to Real Data

Considering that this study aims not only to propose an improved version of a metric designed to assess population QoL but also to validate it, we will include correlation analyses of $QoLI_{v2}$ with GDP and HDI to determine the degree of association between the dependent variable $QoLI_{v2}$ and the two independent variables. It is worth noting that the correlation analysis is limited to the independent variables GDP and HDI, while WHOQOL and BLI are excluded from the calculations for objective reasons.

In the case of WHOQOL, WHO does not publish indicator scores but only offers a QoL measurement methodology. Similarly, OECD does not publish calculated values of BLI, offering only a methodology through which users can determine the ranking of the 38 member countries based on the weighting assigned to each dimension. Furthermore, OECD is an intergovernmental organization comprising a very limited number of countries, and three EU member states (Croatia, Bulgaria, Romania) are not part of this forum, making correlation analysis between $QoLI_{v2}$ and BLI more challenging.

A. *Data Provenance*

The primary source is the QoL database provided by Eurostat [27], which, as the official statistical office of the EU, annually collects, aggregates, and publishes statistical data from both EU member countries and candidate countries seeking accession to the community space. However, since Eurostat does not provide statistics on voter turnout, their scores are sourced from the International Institute for Democracy and Electoral Assistance portal [28]. Regarding GDP values, these are also sourced from statistics provided by Eurostat [29], while the HDI values are extracted from the official UNDP report, *"Human Development Report 2025"* [30].

B. *Correlation Analysis*

To validate the newly introduced formula, Spearman's rank correlation coefficient was calculated between the $QoLI_{v2}$ values presented in Table 2 and two of the most widely used indicators, namely GDP and HDI.

TABLE II. THE VALUES OF THE COMPARED INDICATORS RECORDER BY THE EU MEMBER STATES IN 2023

Country Code	Country Name	GDP per Capita	HDI	$QoLI_{v2}$
AT	Austria	45,510	0.930	30.711
BE	Belgium	44,120	0.951	30.556
BG	Bulgaria	10,970	0.845	29.753
CY	Cyprus	28,670	0.913	30.227
CZ	Czechia	21,660	0.915	30.375
DE	Germany	47,780	0.959	30.587
DK	Denmark	58,640	0.962	30.697
EE	Estonia	21,210	0.905	30.277
EL	Greece	18,670	0.908	29.997
ES	Spain	27,510	0.918	30.363
FI	Finland	43,430	0.948	30.734
FR	France	37,570	0.920	30.406
HR	Croatia	16,060	0.889	30.129
HU	Hungary	16,060	0.870	30.238
IE	Ireland	86,090	0.949	30.560
IT	Italy	32,560	0.915	30.182
LT	Lithuania	19,160	0.895	30.276
LU	Luxembourg	101,450	0.922	30.735
LV	Latvia	17,390	0.889	30.132
MT	Malta	32,740	0.924	30.461
NL	Netherlands	51,010	0.955	30.361
PL	Poland	15,950	0.906	30.390
PT	Portugal	22,020	0.890	30.394
RO	Romania	13,030	0.845	29.966
SE	Sweden	48,510	0.959	30.741
SI	Slovenia	25,050	0.931	30.531
SK	Slovakia	18,750	0.880	30.294

Spearman's rank correlation coefficient, denoted by the Greek letter ρ (rho) or by r_s, is a nonparametric measure of the strength of association between two variables. The values calculated based on this measure range from -1 to 1, with the two ends of the interval representing a perfect relationship, and the value 0 indicating no relationship between the two compared variables.

The results of calculating Spearman's rank correlation coefficient, presented in Figure 1 and Figure 2, indicate a strong monotonic relationship in both cases, with a slight predominance of the correlation with GDP per Capita

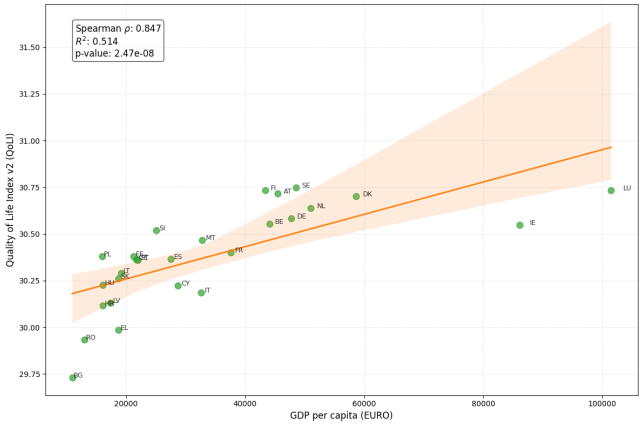


Figure 1. Bivariate plot determined for the variables $QoLI_{v2}$ and GDP per capita (2023).

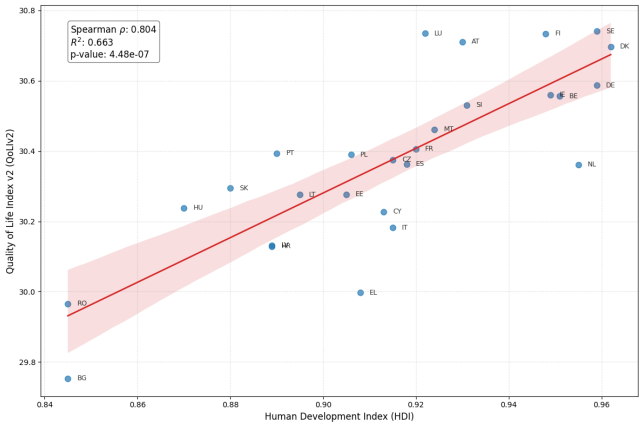


Figure 2. Bivariate plot determined for the variables $QoLI_{v2}$ and HDI (2023).

($\rho = 0.847$, $p < 0.001$) compared to HDI ($\rho = 0.804$, $p < 0.001$), suggesting that the hierarchical positioning of countries is very consistent between economic wealth and the new $QoLI_{v2}$ index.

To reinforce the explanatory power of the indicated relationships, the calculation of R^2 helps us better understand the proportion and variation of $QoLI_{v2}$ explained by the other two variables. Thus, we can observe that 51.40% of the variation in $QoLI_{v2}$ can be attributed to its association with GDP per Capita, while 66.30% of the variation is explained by the correlation with HDI. This difference is largely driven by *economic outliers* like Luxembourg and Ireland, whose exceptional wealth does not translate into a proportionally higher quality of life. This confirms that $QoLI_{v2}$, like HDI, is a multidimensional construct that levels out purely fiscal extremes, offering a more balanced view of social development

IV. Conclusion And Future Work

Although a universally accepted definition for the term QoL has not yet been established, recent years have seen growing interest in examining the indicators that influence population QoL and in developing new metrics to assess it, whether general or specific, related or unrelated to health.

Following a thorough examination of the most well-known metrics and of the techniques used to aggregate indicators, the formula for calculating the QoLI, initially proposed in 2019 [23], has been revisited, and a series of both structural and methodological changes have been introduced. The new version, $QoLI_{v2}$, incorporates a much broader range of atomic indicators than its predecessor (76 indicators compared to 32), significantly expanding the coverage of factors that Eurostat uses in its annual QoL assessments. All these atomic indicators are aggregated into eight objective and one subjective dimension, which are then further aggregated to determine the overall $QoLI_{v2}$ value. In the new formula, aggregation is performed using a logarithmic function instead of the geometric mean.

To validate the new formula, correlation analyses were performed between $QoLI_{v2}$ and two of the most widely recognized indicators: GDP per Capita and HDI. After calculating Spearman's rank correlation coefficient, a strong monotonic correlation was observed in both cases, with values of 0.847 for the relationship between $QoLI_{v2}$ and GDP per Capita, and 0.804 for the relationship between $QoLI_{v2}$ and HDI.

Beyond statistical validation, an important direction for future work involves expanding the interpretative analysis of country-level results. Further investigation is required to substantiate how the rankings produced by $QoLI_{v2}$ reflect socio-economic realities in different national contexts. Comparative case studies between countries exhibiting similar $QoLI_{v2}$ outcomes could provide deeper insight into how distinct economic structures, social policies, or institutional environments lead to comparable QoL results despite differing macroeconomic conditions.

Future research will therefore focus on extending the validation framework through qualitative and comparative analyses, incorporating expert feedback and contextual interpretation of national performances. This includes examining clusters of countries with similar $QoLI_{v2}$ scores, identifying divergence from traditional indicators such as GDP or HDI, and exploring whether $QoLI_{v2}$ captures multidimensional aspects of well-being that remain insufficiently represented in existing development metrics.

References

[1] S. J. Coons, S. Rao, D. L. Keininger, and R. D. Hays, "A comparative review of Generic Quality-of-Life Instruments". Pharmacoeconomics 17(1), pp. 13–35, 2000, doi:10.2165/00019053-200017010-00002.

[2] D. S. J. Costa et al., "How is quality of life defined and assessed in published research?". Qual Life Res 30, pp. 2109–2121, 2021, doi:10.1007/s11136-021-02826-0.

[3] S. M. Hunt, "Editorial: The Problem of Quality of Life". Quality of Life Research 6(3), pp. 205–212, 1997. [Online]. Available from: http://www.jstor.org/stable/4035081 [retrieved: May, 2026].

[4] P. Lagas, F. van Dongen, F. van Rijn, and H. Visser, "Regional quality of living in Europe". REGION 2(2), pp. 1–26, 2015, doi:10.18335/region.v2i2.43.

[5] B. Nováková and V. Šoltés, "Quality of Life Research: Material Living Conditions in the Visegrad Group Countries". Econ. Sociol., 9(1), pp. 282–294, 2016, doi:10.14254/2071-789X.2016/9-1/19.

[6] P. Bartelmus, "Beyond GDP-New approaches to applied statistics". Review of Income and Wealth, 33(4), pp. 347–358, 1987, doi:10.1111/j.1475-4991.1987.tb00679.x.

[7] V. Berenger and A. Verdier-Chouchane, "Multidimensional Measures of Well-Being: standard of living and quality of life across countries". World Development 35(7), pp. 1259–1276, 2007, doi:10.1016/j.worlddev.2006.10.011.

[8] J. van Zanden et al. (eds.), "How Was Life?: Global Well-being since 1820". OECD Publishing, 2014, doi:10.1787/9789264214262-en.

[9] E. Diener and E. Suh, "Meaasuring quality of life: economic, social and subjective indicators". Social Indicators Research 40, pp. 189–216, 1997, doi:10.1023/A:1006859511756.

[10] H. C. Gift and M. Redford, "Oral Health and The Quality Of Life". Clinics in Geriatric Medicine 8(3), pp. 673–684, 1992, doi:10.1016/S0749-0690(18)30471-3.

[11] D. Bennadi and C. V. K. Reddy, "Oral health related quality of life". Journal of International Society of Preventive and Community Dentistry 3(1), pp. 1-6, 2013, doi:10.4103/2231-0762.115700.

[12] R. M. Baiju, E. Peter, N. O. Varghese, and R. Sivaram, "Oral health and quality of life: current concepts". J Clin Diagn Res. 11(6), pp. ZE21–ZE26, 2017, doi:10.7860/JCDR/2017/25866.10110.

[13] M. G. Nayak et al., "Quality of Life among Cancer Patients". Indian J Palliat Care 23(4), pp. 445-450, 2017, doi:10.4103/IJPC.IJPC_82_17.

[14] M. Krawczyk-Suszek and A. Kleinrok, "Health-Related Quality of Life (HRQoL) of People over 65 Years of Age". Int. J. Environ. Res. Public Health 19(2), pp. 625, 2022, doi:10.3390/ijerph19020625.

[15] S. Noto, "Perspectives on Aging and Quality of Life". Healthcare 11(15), pp. 2131, 2023, doi:10.3390/healthcare11152131.

[16] K. Kim, M. Uysal, and M. J. Sirgy, "How does tourism in a community impact the quality of life of community residents?". Tourism Management 36, pp. 527–540, 2013, doi:10.1016/j.tourman.2012.09.005.

[17] H. Ramkissoon, "Perceived social impacts of tourism and quality-of-life: a new conceptual model". Journal of Sustainable Tourism 31(2), pp. 442–459, 2020, doi:10.1080/09669582.2020.1858091.

[18] J. M. Shapiro, "Smart Cities: Quality of Life, Productivity, and the Growth Effects of Human Capital". The Review of Economics and Statistics 88 (2), pp. 324–335, 2006, doi:10.1162/rest.88.2.324.

[19] J. C. F. De Guimarães, E. A. Severo, L. A. Felix Júnior, W. P. L. B. Da Costa, and F. T. Salmoria, “Governance and quality of life in smart cities: Towards sustainable development goals”. Journal of Cleaner Production, 253, 2020, doi:10.1016/j.jclepro.2019.119926.

[20] E. Neumayer, “The human development index and sustainability – a constructive proposal”. Ecological Economics 39(1), pp. 101–114, 2001, doi:10.1016/S0921-8009(01)00201-4.

[21] UNDP Human Development Report Office, Training Material for Producing National Human Development Reports, 2015. [Online]. Available from: https://hdr.undp.org/system/files/documents/hditraining.pdf [retrieved: May, 2026].

[22] The WHOQOL Group: The World Health Organization quality of life assessment (WHOQOL): Position paper from the World Health Organization. Social Science & Medicine 41(10), pp. 1403–1409, 1995, doi:10.1016/0277-9536(95)00112-K.

[23] I. C. Dorobăț, O. Rinciog, G. C. Muraru, and V. Posea, “Quality of Life Index Analysis for the Case of Romanian Regions”. Proceedings of The Thirteenth International Conference on Digital Society and eGovernments (ICDS 2019) IARIA, 2019, pp. 37–44, ISSN: 2308-3956, ISBN: 978-1-61208-685-9.

[24] I. C. Dorobăț and V. Posea, “eLIF: European Life Index Framework - An Analysis for the Case of European Union Countries”. International Journal on Systems and Measurements 12(3&4), pp. 198–214, 2019, ISSN: 1942-261x.

[25] J. M. Bland and D. G. Altman, “The Use Of Transformation When Comparing Two Means”. BMJ 312(7039), pp. 115, 1996, doi:10.1136/bmj.312.7039.1153.

[26] Your Europe. Teenage workers in the EU: Age limits & working time. [Online]. Available from: https://europa.eu/youreurope/business/human-resources/employment-contracts/teenage-workers/index_en.htm [retrieved: May, 2026].

[27] Quality of Life – Database, Eurostat. [Online]. Available from: https://ec.europa.eu/eurostat/web/quality-of-life/database [retrieved: May, 2026].

[28] International IDEA Home Page. [Online]. Available from: https://www.idea.int/ [retrieved: May, 2026].

[29] Data Browser – Real GDP per Capita, Eurostat. [Online]. Available from: https://ec.europa.eu/eurostat/databrowser/view/sdg_08_10/default/table?lang=en [retrieved: May, 2026].

[30] UNDP Human Development Report Office, Human Development Report 2025, 2025. [Online]. Available from: https://hdr.undp.org/content/human-development-report-2025 [retrieved: May, 2026].

Designing a Data-Driven Decision Support System for Sustainable and Climate-Resilient Forest Management: Lessons Learned from the OptForEU Project

Lasse Berntzen
School of Business
University of South-Eastern Norway
Hønefoss, Norway
e-mail: lasse.berntzen@usn.no

Marius Rohde Johannessen
School of Business
University of South-Eastern Norway
Borre, Norway
e-mail: marius.johannessen@usn.no

Alessio Collalti
Forest Modelling Laboratory
National Research Council of Italy
Perugia, Italy
e-mail: alessio.collalti@cnr.it

Mauro Morichetti
Forest Modelling Laboratory
National Research Council of Italy
Perugia, Italy
e-mail: mauro.morichetti@cnr.it

Hermine Mitter
Department of Environmental Systems Sciences
University of Graz
Graz, Austria
e-mail: hermine.mitter@uni-graz.at

Stefanie Linser
Institute of Forest, Environmental and Natural Resource Policy
BOKU University
Vienna, Austria
e-mail: stefanie.linser@boku.ac.at

Alice Ludvig
Department of Economic and Social Sciences
BOKU University
Vienna, Austria
e-mail: alice-ludvig@boku.ac.at

Francesca Gianetti
Department of Agriculture, Food, Environment and Forestry
University of Florence
Florence, Italy
e-mail: francesca.giannetti@unifi.it

Ilaria Zorzi
Bluebiloba Startup Innovativa s.r.l
Florence, Italy
e-mail: ilaria.zorzi@bluebiloba.com

Sorin Cheval
National Meteorological Administration
Bucharest, Romania
e-mail: sorin.cheval@meteoromania.ro

***Abstract*—Climate change mitigation and adaptation increasingly rely on effective forest management, as forests represent one of Europe's largest carbon sinks while simultaneously providing biodiversity, economic, and social benefits. However, forest management decisions are complex due to uncertain climate futures, diverse ecosystem services, and regionally varying management practices. The OptForEU project addresses this challenge by developing a data-driven Decision Support System (DSS) to assist forest managers and policymakers in evaluating management strategies across different climate and policy scenarios. While most project outputs target forestry and environmental science audiences, this paper presents the project from an Information Systems (IS) perspective. The paper describes the background, design principles, data architecture, and implementation of the DSS, emphasizing the challenges of integrating heterogeneous data sources, fostering cross-disciplinary collaboration, and involving stakeholders in co-creation. Lessons learned from large-scale European collaboration are discussed, with particular attention to user adoption, ontology mismatches, and the socio-technical challenges of deploying data-driven decision tools across diverse national contexts. The paper concludes with reflections on how IS methods and co-creation practices can improve the adoption of environmental decision-support tools in complex multi-stakeholder settings.**

Keywords-forest management; decision support; co-creation; open data; OptForEU.

I. INTRODUCTION

The OptForEU project has produced numerous scientific outputs, primarily for forestry researchers, climate scientists, and environmental policymakers. However, the project also represents a substantial effort in building an operational, data-driven Decision Support System (DSS), raising important Information Systems and socio-technical challenges that are highly relevant beyond forestry research, particularly for forest owners and forest managers. This paper, therefore, presents the project from an Information Systems perspective, focusing on system design, data integration, stakeholder collaboration, and implementation lessons relevant to IT professionals and IS researchers.

Forests play a central role in climate mitigation because they store and sequester large amounts of carbon while simultaneously providing biodiversity, recreation, and economic value through timber production and other ecosystem services. At the same time, climate change, induced biotic and abiotic disturbances (e.g., droughts, pest outbreaks, forest fires, windstorms), and evolving economic and policy demands create complex trade-offs between conservation objectives and resource utilization. Forest management decisions must balance unmanaged or old-growth forests, which often maximize biodiversity and long-term carbon storage, with sustainable and integrated forest management that supplies renewable materials and supports

economic activity while still contributing to ecosystem services.

Forest management decisions differ from many other policy and management decisions in that their consequences unfold over decades and are often irreversible. Choices regarding harvesting intensity, species composition, or conservation set forest trajectories that constrain future options and lock in ecological and economic outcomes for generations [1].

In practice, forest management decisions are often based on a combination of expert judgment, locally developed planning tools, and fragmented datasets. While such approaches may work within stable contexts, they offer limited support for systematically evaluating long-term trade-offs under uncertain climate futures or comparing alternative strategies across regions [2].

While DSSs have been studied for decades in environmental and natural resource management, the literature has often prioritized model development and technical performance. Less attention has been given to the socio-technical conditions of implementation, including stakeholder participation and co-creation, trust, institutional fit, and cross-context adaptation in heterogeneous policy and cultural settings [3].

Decision-makers therefore face long-term choices under uncertain climate and economic scenarios, often with incomplete information and regionally specific constraints.

The OptForEU project seeks to address this complexity by providing decision-support tools that enable stakeholders to evaluate alternative management strategies under varying environmental and policy assumptions. This paper explores why a DSS approach is needed in forest management, how data-driven methods support such decision-making, how stakeholder co-creation shapes system design, and which lessons can be learned from cross-disciplinary system development.

Through this, the paper makes three main contributions: First, the paper provides a detailed account of the design and architecture of a large-scale, data-driven DSS for forest management developed in a European research context. Second, it identifies key socio-technical challenges related to data integration, stakeholder trust, and cross-national collaboration. Third, it contributes to Information Systems research by reflecting on how co-creation and design science approaches can enhance the adoption of environmental decision-support tools. The paper is guided by the following questions:

1. *How can a data-driven DSS support forest management decisions under climate and policy uncertainty?*
2. *What socio-technical challenges emerge when such systems are co-created with diverse stakeholders?*
3. *What lessons can be drawn for the design and adoption of environmental DSS in complex governance contexts?*

Section II describes the design process. Section III presents the data-driven decision framework. Section IV focuses on co-creation and stakeholder-centered design. Section V presents results and lessons learned. Section VI discusses the findings. Section VII concludes the paper and proposes future work.

II. Designing a DSS for Forest Management

DSSs have long been used in domains characterized by uncertainty, multiple competing objectives, and incomplete information. Forest management represents precisely such a domain, where decisions often affect ecosystems and economic conditions for decades or even centuries.

Forest management decisions must simultaneously consider carbon sequestration capacity, biodiversity preservation, timber production, management intensity, climate change adaptation, soil conservation and health, economic viability, and regional policy constraints. Traditionally, such decisions rely heavily on expert judgment and locally developed practices, making systematic comparisons across alternative forest management practices difficult. The OptForEU DSS aims to address this challenge by providing tools that simulate different forest management practices under three different climate change scenarios, quantify the impacts of ecosystem services, and support transparent decision-making processes, all based on a set of Essential Forest Management Indicators (EFMIs).

From an Information Systems perspective, the system must achieve analytical accuracy while remaining usable and trustworthy for practitioners. Results must be interpretable, transparent, and aligned with existing decision-making practices. Major design challenges include accommodating diverse user groups across different European regions and countries, handling heterogeneous data sources, maintaining transparency in modeling outputs, and ensuring adaptability across different forest types and governance contexts.

III. Data-Driven Decision Framework

The OptForEU DSS relies on integrating a wide variety of data sources into a unified analytical framework. Central inputs include satellite imagery (e.g., European Copernicus data), national forest inventory data, climate projections, soil and biodiversity datasets, model output, open data repositories, and records of forest management practices.

While empirical and observational data describe current and historical forest conditions, they cannot directly provide projections of future forest development under alternative climate and management scenarios. Model-generated datasets are therefore used to support long-term scenario analysis under climate uncertainty.

The DSS is informed by a suite of model-generated datasets describing forest dynamics under alternative climate and management scenarios. In particular, outputs from forest ecosystem models, including 3D-CMCC-FEM [4] and PICUS [5], provide projections of EFMIs and information on forest growth, carbon fluxes, and structural changes. These are complemented by regional climate model outputs (e.g., RegCM-CCLM [6] and REMO-iMOVE [7]), and by land-surface model simulations (e.g., JULES [8]), which describe land–atmosphere exchanges and surface processes. Together, these modeled datasets form the core knowledge base used to "fill" the DSS and support the evaluation of management options across varying climatic and policy contexts.

Each modeling component contributes complementary information to the DSS. Forest ecosystem models provide management-sensitive indicators, including productivity, carbon sequestration potential, mortality risk, and structural development under alternative silvicultural strategies. Regional climate models supply spatially explicit projections of temperature, precipitation, and climate extremes, enabling the assessment of exposure to future climatic conditions and other EFMIs. Land-surface model outputs further enrich the dataset by describing hydrological and energy exchanges, which are relevant for evaluating ecosystem functioning and climate-related feedback.

To reduce overlap and inconsistencies, outputs are harmonized into a common indicator framework, aligned spatial formats, and compatible temporal resolutions before integration into the DSS. This allows comparisons across models, regions, and scenarios despite differences in spatial resolution, temporal coverage, and variable definitions. The use of multiple models also improves robustness by enabling ensemble-based analysis across different assumptions and scenarios.

By integrating multiple models and scenarios, the DSS relies on an ensemble-based knowledge base that captures a range of plausible futures. This approach allows users to explore trade-offs among management strategies and assess their robustness under varying climatic and environmental conditions. In this way, the DSS leverages process-based modeling and climate projections to provide a transparent and data-driven foundation for decision-making in complex forest management contexts.

Integrating these sources poses substantial technical challenges. Data must be normalized across regions that use different measurement standards and collection methodologies, while differences in temporal resolution complicate time-series analysis. In addition, missing or incomplete data must be handled carefully to avoid misleading conclusions, and large-scale spatial datasets create significant computational demands.

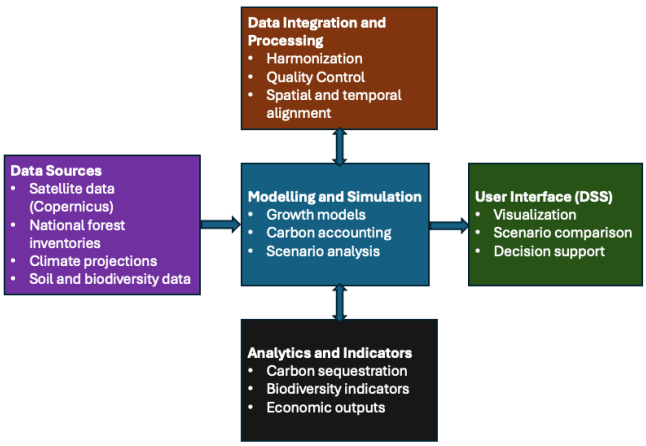


Figure 1. Conceptual architecture of the OptForEU DSS.

For an IT-oriented audience, this part of the project highlights data engineering and integration challenges rather than domain-specific forestry science. The project demonstrates how open data and Earth observation technologies enable scalable environmental decision support, while also revealing persistent challenges related to data quality, harmonization, and computational performance.

As illustrated in Figure 1, the OptForEU DSS follows a layered architecture separating data acquisition, integration, modeling, and user interaction. This separation supports scalability across regions and allows individual components, such as climate scenarios or management models, to be updated independently. The architecture illustrates the flow of heterogeneous environmental and management data sources through data integration and modeling layers to user-facing decision-support functionalities.

The layered structure emphasizes data harmonization, scenario-based simulation, and indicator-based analytics supporting transparent and comparable forest management decisions across regions.

IV. Co-Creation and Stakeholder-Centered Design

The development of the OptForEU DSS revealed that technical accuracy alone is insufficient for creating decision-support tools that will be trusted, adopted, and integrated into real-world forest management practice. Because forest governance involves multiple stakeholders with different objectives, regulatory responsibilities, professional traditions, and epistemic cultures, the project adopted a co-creation approach to ensure that the DSS would be relevant, usable, and legitimate across diverse national contexts. The adoption of co-creation as a design strategy reflects long-standing traditions in participatory design that emphasize the involvement of users and practitioners in shaping technological artifacts [9][10]. Co-creation in this setting goes beyond consultation: it is an iterative socio-technical design process in which stakeholders and system developers jointly shape system requirements, modeling assumptions, and user-facing functionalities (ibid.).

Forest managers, policymakers, scientists, NGOs, and environmental agencies often operate with divergent priorities, including biodiversity protection, timber production, climate mitigation, economic viability, and cultural or recreational values. These priorities not only differ across groups but also across countries, making it unlikely that a one-size-fits-all DSS would align with existing decision-making practices. In addition, many forest management decisions rely heavily on experiential knowledge and local practices, which are not easily captured in technical models or harmonized datasets. Co-creation was therefore essential for bridging these epistemic and institutional differences, supporting mutual understanding, and securing stakeholder buy-in for the joint indicator development, the modeling framework, and the system outputs. Participatory design research shows that involving stakeholders early is essential for developing artifacts that are legitimate and meaningful to practitioners [11].

A. Co-creation methods and process

The OptForEU co-creation process used a combination of multi-stakeholder workshops, semi-structured interviews, prototype testing sessions, and iterative feedback loops. These methods mirror participatory design approaches that promote collaborative problem articulation and design refinement [10]. Workshops held in partner countries and in national languages brought together forest managers, policymakers, scientists, and environmental organizations. Early prototypes—ranging from conceptual models to interactive interface mockups—were presented to stakeholders, enabling evaluation of system transparency, interpretability, and relevance.

Feedback was synthesized through thematic analysis and systematically integrated into subsequent design iterations, consistent with Design Science Research principles emphasizing iterative evaluation and refinement [12][13]. This cyclical process not only improved system usability but also helped uncover implicit knowledge and assumptions held by stakeholders.

B. Cross-national and cross-disciplinary challenges

The co-creation process highlighted several socio-technical challenges inherent in developing a DSS applicable from the local to the pan-European level. First, stakeholders used different terminologies and conceptual framings for forest conditions, risks, and management practices. Terms such as "old-growth forest," "resilience," or "ecosystem services" carried different meanings across countries and professional groups, necessitating shared definitions to avoid ambiguity in model outputs and indicators.

Second, regulatory diversity across Europe affected how users interpreted the relevance of certain scenarios or indicators. What is considered a realistic management alternative in one country may be legally restricted or culturally unacceptable in another. Differences in terminology and institutional practices can hinder useful collaboration and adoption unless shared representations and understandings are developed [14].

Third, differences in digital maturity and model literacy created asymmetries in stakeholders' ability to evaluate and critique early prototypes. Some users preferred fine-grained, stand-level outputs supported by national inventory data, while others emphasized regional scenario comparisons or policy-level summaries. Navigating these differences required a flexible design approach that could accommodate varying levels of detail and analytical complexity without compromising transparency or usability. Post-implementation, the project will distribute a survey to pilot testers to evaluate the perceived usefulness of the DSS, based on technology acceptance literature [15][16].

C. Socio-technical design tensions

A central finding from the co-creation process was the presence of socio-technical design tensions that shaped the system architecture. One such tension concerned the trade-off between model sophistication and interpretability. Stakeholders consistently emphasized the need for transparent models and easily comprehensible indicators, even if this required simplifying some components. Another tension involved balancing harmonized, European-wide data structures with the need for local relevance. While harmonization supports comparability, it risks obscuring important local conditions or management practices, leading to skepticism among practitioners.

A third tension emerged between users' desire for flexibility in creating custom scenarios and the need to maintain analytical rigor and prevent misuse or misinterpretation of model outputs. Addressing these tensions required continuous negotiation between system developers and domain experts, underscoring the socio-technical nature of environmental DSS development. This aligns with findings in the socio-technical literature, which emphasize that trust in systems emerges from transparency, accountability, and comprehensibility [17][18].

D. Co-creation recommendations

Overall, co-creation activities have provided valuable input to the DSS design team, emphasizing the importance of reducing dashboard complexity, providing guidance on the EFMIs, which are strongly related to forest ecosystem services [19], and providing feedback on how output should be structured to fit with relevant policies.

Based on these observations, the OptFor-EU project recommends the following for future EU-funded projects:

At the *policy level*, establish co-creation as a core principle in multidisciplinary projects that solve real-world, local-to-higher-level problems. This should be an incentivized component of EU-funded projects, where applicable. Unless co-creation is embedded throughout work packages, it can easily become secondary to other objectives.

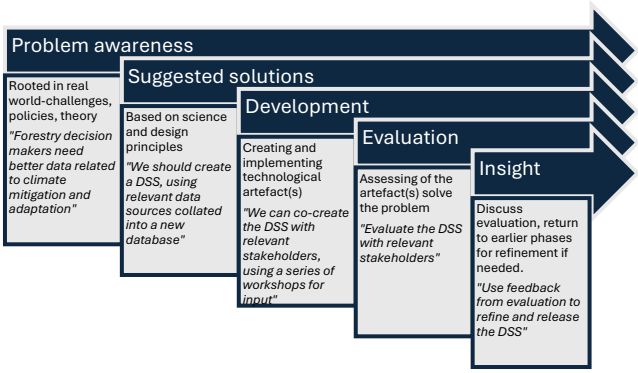


Figure 2. Co-creation design science process. Based on [12].

The principle of co-creation requires clear and comprehensive guidelines and best practices for successful integration into projects. It is easy to think of co-creation as an additional activity, but for it to succeed, it must be a core activity to which every project partner is committed.

Building on the previous point, support capacity building by investing in EU-wide training programs for project teams and stakeholders on co-creation and design thinking.

At the *practical level,* projects aimed at creating a digital artifact should also use design thinking as their methodology and consider employing design science research [12] as their research methodology. Resources for co-creation activities

need to be allocated in the budgeting process. The project timeline should include activities, such as design thinking workshops for insight and ideation, stakeholder workshops for idea development and user feedback, and the evaluation of the artifact. Figure 2 illustrates this process.

V. Results and Lessons Learned

The project successfully produced an operational DSS prototype capable of simulating forest management scenarios across diverse European contexts. The project delivered an integrated modeling framework, a functional DSS platform, stakeholder engagement mechanisms, and validated case studies demonstrating applicability across different forest regions.

However, long-term success ultimately depends on adoption beyond the project itself. Several lessons emerged during development. Interdisciplinary collaboration required significant effort to align terminology and conceptual frameworks among partners from forestry, climate science, and Information Systems. Data harmonization consumed considerably more time and resources than initially expected. Building user trust required sustained engagement, transparency, and demonstration of system value. The project also demonstrated that Information Systems expertise plays a crucial role in environmental digitalization initiatives, as technical integration and adoption challenges often dominate implementation success.

Overall, the project illustrates that technological capability alone does not guarantee adoption; socio-technical considerations remain equally important.

VI. Discussion

This paper set out to explore how data-driven DSSs can support forest management under uncertainty, which socio-technical challenges emerge in their development, and what lessons can be drawn for designing and implementing such systems in complex governance contexts. Based on the OptForEU project, several key insights emerge.

A. Supporting forest management under uncertainty

The findings demonstrate that a data-driven DSS can significantly enhance forest management decision-making by enabling the systematic evaluation of alternative strategies across multiple dimensions and time horizons. By integrating heterogeneous data sources—including climate projections, forest ecosystem models, and environmental indicators—the DSS provides a structured way to explore trade-offs between carbon sequestration, biodiversity, and economic outcomes.

Rather than replacing expert judgment, the DSS complements existing decision-making practices by making assumptions explicit, enabling scenario comparison, and improving transparency. The use of ensemble-based modeling further allows decision-makers to assess the robustness of management strategies under varying climate and policy conditions. In this way, the DSS reduces uncertainty not by eliminating it, but by making it visible, structured, and analyzable.

B. Socio-technical challenges in co-created DSS

The development of the OptForEU DSS highlights that the primary challenges in environmental decision-support systems are not purely technical but socio-technical. Three recurring design tensions were particularly evident.

First, a tension between model sophistication and interpretability emerged. While advanced models increase analytical accuracy, stakeholders consistently prioritized transparency and comprehensibility. This required simplifying certain outputs and focusing on interpretable indicators rather than maximizing model complexity.

Second, a tension between harmonization and local relevance became apparent. The need to integrate data across European regions necessitated standardized indicators and data structures. However, this standardization risked obscuring local conditions, practices, and regulatory constraints, which are critical for practitioner trust and adoption.

Third, a tension between flexibility and analytical rigor was observed. Stakeholders expressed interest in customizing scenarios and exploring alternative assumptions, but unrestricted flexibility can lead to misuse or misinterpretation of model outputs. Balancing user autonomy with methodological robustness required careful interface and system design.

In addition to these tensions, the co-creation process revealed challenges related to differences in terminology, varying levels of digital maturity, and institutional diversity across countries. These factors significantly influenced how stakeholders interpreted system outputs and engaged with the DSS.

C. Lessons for design and adoption of environmental DSS

The OptForEU experience provides several broader lessons for Information Systems research and practice.

First, co-creation is essential but must be treated as a core design activity rather than a supplementary process. Continuous stakeholder involvement not only improves usability but also builds trust, aligns expectations, and uncovers implicit domain knowledge that cannot be captured through technical modeling alone.

Second, data integration is a central challenge that often exceeds initial expectations. Harmonizing heterogeneous datasets across spatial, temporal, and conceptual dimensions requires substantial effort and should be explicitly accounted for in system design and project planning.

Third, trust and adoption depend on transparency and interpretability as much as on analytical capability. Users are more likely to adopt systems that clearly communicate assumptions, limitations, and uncertainties, even if this comes at the cost of reduced model complexity.

Fourth, flexible system architectures are necessary to accommodate diverse user needs and governance contexts. Layered and modular designs enable adaptation to different regions and policy environments while maintaining a consistent analytical core.

Finally, the findings reinforce the importance of a socio-technical perspective in digital transformation initiatives. Successful deployment of DSS solutions depends not only

on technical performance but also on alignment with institutional practices, stakeholder expectations, and governance structures.

D. Contributions

This paper contributes to Information Systems research in three ways. First, it provides an empirical account of designing and implementing a large-scale, data-driven DSS in a complex, multi-stakeholder environmental domain. Second, it identifies key socio-technical design tensions that extend existing literature on DSS and participatory design. Third, it demonstrates how co-creation and design science approaches can improve the adoption and legitimacy of decision-support tools in real-world settings.

VII. Conclusion and Future Work

The OptForEU project illustrates both the potential and the complexity of applying digital technologies to environmental decision-making. A data-driven DSS can provide valuable support for climate-resilient forest management by enabling structured exploration of uncertainty and trade-offs. However, the success of such systems depends as much on socio-technical factors as on technical design.

For Information Systems researchers and practitioners, environmental DSS represents an important and growing application domain where digital transformation directly intersects with sustainability and climate policy. Addressing the socio-technical challenges identified in this study will be essential for ensuring that such systems are not only technically robust but also trusted, adopted, and impactful in practice.

Future research should further investigate methods for balancing model complexity and usability, explore scalable approaches to co-creation in large international projects, and examine long-term adoption and impact of DSS tools in operational forest management. In addition, there is a need to develop evaluation frameworks that capture not only technical performance but also trust, legitimacy, and institutional fit.

Acknowledgment

This study was conducted within and funded by the project “OPTimising FORest management decisions for a low-carbon, climate resilient future in Europe (OptFor-EU)” funded by the European Union Horizon Europe programme, under Grant Agreement n°101060554.

References

[1] J. Abildtrup, J. Laye, M. Laye, and A. Stenger, "Irreversibility and uncertainty in multifunctional forest management allocation," in Global Perspectives on Sustainable Forest Management, O. C. Akais Ed. InTech, pp. 263-274, 2012.

[2] T. McDaniels, T. Mills, R. Gregory, and D. Ohlson, “Using expert judgments to explore robust alternatives for forest management under climate change,” Risk Analysis, vol. 32(12), pp. 2098-2112, 2012, doi: 10.1111/j.1539-6924.2012.01822.x.

[3] E. Walling and C. Vaneeckhaute, “Developing successful environmental decision support systems: Challenges and best practices,” Journal of Environmental Management, vol. 264, 110513, 2020.

[4] A. Collalti et al., “Thinning can reduce losses in carbon use efficiency and carbon stocks in managed forests under warmer climate,” Journal of Advances in Modeling Earth Systems, vol. 10(10), pp. 2427-2452, 2018, doi: 10.1029/2018MS001275.

[5] F. Irauschek et al., “Evaluating five forest models using multi-decadal inventory data from mountain forests,” Ecological Modelling, vol. 445, 109493, pp. 1-11, 2021, doi: 10.1016/j.ecolmodel.2021.109493.

[6] F. Giorgi et al., "RegCM4: model description and preliminary tests over multiple CORDEX domains," Climate research, vol. 52, pp. 7-29, 2012, https://doi.org/10.3354/cr01018

[7] C. Asmus, P. Hoffmann, J-P. Pietkäinen, J. Böhner, and D. Rechid, "Modeling and evaluating the effects of irrigation on land–atmosphere interaction in southwestern Europe with the regional climate model REMO2020–iMOVE using a newly developed parameterization," Geoscientific Model Development, vol. 16(24), pp. 7311-7337, 2023, https://doi.org/10.5194/gmd-16-7311-2023

[8] M. J. Best et al. "The Joint UK Land Environment Simulator (JULES), model description–Part 1: energy and water fluxes," Geoscientific Model Development, vol. 4(3), pp. 677-699, 2011. https://doi.org/10.5194/gmd-4-677-2011.

[9] D. Schuler and A. Namioka (Eds.), Participatory design: Principles and practices. Boca Raton: CRC Press. 1993.

[10] S. Bødker, F. Kensing, and J. Simonsen, Participatory IT design: Designing for business and workplace realities. Cambridge, MA: MIT Press. 2004.

[11] C. Spinuzzi, “The methodology of participatory design,” Technical Communication, vol. 52(2), pp. 163–174, 2005.

[12] A. R. Hevner, S. T. March, J. Park, and S. Ram, “Design science in information systems research,” MIS Quarterly, vol. 28(1), pp. 75–105, 2004.

[13] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, “A design science research methodology for information systems research,” Journal of Management Information Systems, vol. 24(3), pp. 45–77, 2007, https://doi.org/10.2753/MIS0742-1222240302.

[14] P. R. Carlile, “A pragmatic view of knowledge and boundaries: Boundary objects in new product development,” Organization Science, vol. 13(4), pp. 442–455, 2002.

[15] F. D. Davis, “Perceived usefulness, perceived ease of use, and user acceptance of information technology,” MIS Quarterly, vol. 13(3), pp. 319–340, 1989.

[16] V. Venkatesh, M. G. Morris, G. B. Davis, and F. Davis, “User acceptance of information technology: Toward a unified view,” MIS Quarterly, vol. 27(3), pp. 425–478, 2003.

[17] R. C. Mayer, J. H. Davis, and F. D. Schoorman, “An integrative model of organizational trust,” Academy of Management Review, vol. 20(3), pp. 709–734, 1995.

[18] G. Baxter and I. Sommerville, “Socio-technical systems: From design methods to systems engineering,” Interacting with Computers, vol. 23(1), pp. 4–17, 2011.

[19] S. Linser et al., Report on a novel set of Essential Forest Mitigation Indicators (EFMIs). OptForEU deliverable 1.2. [Online]. Available from: https://optforeu.eu/wp-content/uploads/2025/04/OptFor-EU_D1.2_-EFMIs-_v02_20250314_BOKU.pdf [Retrieved: 31.03.2026]

Will GenAI Make or Break Your Process? - Structuring the Influence of GenAI on Business Process Resilience

Olga Levina
Brandenburg University of Applied Sciences
Brandenburg an der Havel, Germany
e-mail: levina@th-brandenburg.de

Abstract— **The swift adoption of Generative Artificial Intelligence (GenAI) in organizations prompts questions about its effects on Business Process Resilience (BPR). While GenAI is often linked to increased productivity, its influence on the robustness and stability of existing processes remains poorly understood. This research explores how GenAI integrates into established business processes, viewing it as a disruptor that challenges, rather than strengthens, process resilience. A combination of qualitative analysis and process simulation indicates that traditional performance metrics are insufficient for evaluating GenAI's impacts. Instead, explicitly assessing process and output quality, as well as human involvement, is necessary to understand new trade-offs. The paper points to human–organizational challenges, such as increased workload, technostress, and evolving roles for human process actors, which will require a focus on evaluation, oversight, and validation. The qualitative and the simulation-based approaches to assessing the impact of GenAI on BPR provide a diverse assessment of its impacts. These insights suggest that the introduction of GenAI into a business process should be managed as a socio-technical intervention.**

Keywords-Generative AI; business process resilience; business process performance; simulation; systemic view.

I. INTRODUCTION

Organizations operate in dynamic, ever-changing environments, where disruptions such as surges in case arrivals, equipment breakdowns, or workforce absences are common [1]. These incidents can damage personnel or equipment and interrupt operations. The first crucial step for organizations aiming to reduce the negative impact of these disruptions is to evaluate the resilience of their processes [2]. Resilience helps prepare for and manage adverse events by focusing on the resources needed to sustain processes, thereby helping control their impacts. Understanding these dynamics helps manage disruptive events and their negative effects on enterprise operations and resources. This concept also offers a systemic approach to addressing these challenges, encompassing the capacity to prepare for, prevent, protect against, respond to, and recover from setbacks [3]. It encompasses attributes such as robustness [4], adaptability [5], and redundancy [6].

Software tools based on Generative Artificial Intelligence (GenAI) are often integrated into live processes top-down without additional employee training [7], risking cybersecurity risks, skill gaps, and stakeholder issues [8]. For business process actors, this introduction occurs as an immediate and potentially job-threatening event. The immediate introduction has the disadvantage that specific GenAI tasks are not defined within the existing process flow, their applications are not managed, and the role of their outputs is not discussed. Hence, the introduction of GenAI into a current, non-redesigned business process is considered an adversarial event in this research and a challenge for Business Process Resilience (BPR).

Using BPR characteristics from [1] such as absorption, adaptability, agility, redundancy, robustness, recovery, resourcefulness, and rapidity, this research addresses the research question: What is the potential impact of GenAI introduction on BPR? A structured analysis maps existing findings onto BPR characteristics. A simulation of a workflow before and after GenAI highlights its effects, especially challenging aspects like absorption, adaptability, and flexibility. The combined qualitative and simulation approach offers a multidimensional assessment of technology's impact on processes, resources, and control flow—helping researchers study BPR effects and guiding managers to avoid compromising resilience and utilize GenAI effectively.

The paper is organized as follows: Section 2 discusses business process resilience and dimensions; Section 3 evaluates GenAI's impact; an exemplary process is analyzed through qualitative assessment and simulation; and finally, the paper concludes with discussions and future outlook.

II. BUSINESS PROCESS RESILIENCE (BPR)

The concept of resilience is present in different disciplines. An overarching understanding is provided by Haimes [9]. The author describes resilience as a system's ability to withstand disruption with acceptable degradation and to recover within a suitable time and at reasonable cost. Furthermore, the resilience literature distinguishes between a system's bouncing back and bouncing forward after a disruption, and frames bouncing forward as a sign of a new competitive advantage [10]. Duchek [11] provides a widely accepted definition, adopted here, that captures both reactive and proactive views, defining resilience as the ability to anticipate threats, cope with unexpected events, and learn from them.

There has been limited research assessing resilience at the process level. Kraus et al. [1] present a reactive view of

process resilience as "an organization's ability to restore a process to its acceptable performance level after a disruption." Here, the definition focuses on process characteristics and is extended by Duchek's [11] general view. Process resilience is defined here as the ability of the process to anticipate threats, cope with unexpected events, and restore its performance to an acceptable level after disruption.

The topic of measuring or monitoring process resilience has not yet attracted notable interest in the information systems community. Zahoransky et al. [12] present a framework for detecting process resilience properties through log file analysis. Bhuiyan et al. [13] consider process actors to be critical and vulnerable in the context of process resilience. This approach focuses on enabling managers and analysts to manage human resources in ways that maintain process resilience, i.e., by delegating dependencies among actors, choosing alternatives, decomposing tasks, maintaining consistency between organizational and process models, or handling exceptions. The approach by Lee et al. [14] combines the two views. It uses process mining and social network analysis to define a metric called the "degree of substitution," which measures the extent to which the work experiences of human resources overlap, considering two perspectives: task execution and work transfer. It uses event logs for these analyses. Hence, human resources are treated as interchangeable elements in the resilience process.
Here, business process actors are considered integral to the process and, thus, to process resilience. Hence, in addition to the process performance metrics and process logic captured in the process model, business process resilience is considered here in the context of business process actors.

Kraus et al. [1] identify the following characteristics of business process resilience: absorption, adaptability, agility, flexibility, redundancy, robustness, recovery, resourcefulness, and rapidity. *Absorption* is described by [1] as the ability to dampen the impact of disruptive events. Kule [15] introduces, in this context, the additional characteristic of a system's absorptive capacity. It is defined as an organization's ability to acquire, assimilate, transform, and exploit knowledge, a capability that has emerged as a critical enabler of resilience. The ability to recognize the value of new knowledge is a critical yet subjective component of absorptive capacity [15]. *Adaptability* refers to the ability to respond to and adjust to a changing environment. Duchek [11] extends the term with system's adaptive capacity. It indicates an organization's ability to handle disruptions and bounce back, highlighting flexibility and the reorganization of resources [11]. *Agility* refers to the ability to quickly respond to shifting conditions. It involves sensing and adapting to changes in the business environment that could affect production processes, thereby enabling faster responses [16]. The process includes monitoring indicators and taking action to implement necessary changes.

Flexibility refers to the system's ability to change and adapt to new or complex situations. It also includes the ability to incorporate alternative execution paths during design time, affecting either the process instance or the overall business process model [17]. *Redundancy* involves duplicate processes, resource allocation, and the additional capacity to withstand potentially severe disruptions. *Robustness* refers to the capability to endure a certain level of stress without significant loss of function. It often involves maintaining critical operations by mobilizing resources, activating contingency plans, and making quick decisions, sometimes including temporary downsizing [10]. *Recovery* is the ability to quickly restart operations and reach a targeted performance level. *Resourcefulness*: process monitoring and controlling; The ability to diagnose problems and to initiate solutions. *Rapidity*: The ability to react fast to changes in its environment.

In this research, these characteristics are used to analyse and structure the impact of GenAI introduction into a business process on the process resilience.

III. IMPACT OF GENAI INTRODUCTION ON BPR

Generative AI (GenAI) tools, defined as any "end user tool [...] whose technical implementation includes a generative model based on deep learning" [18], are the latest in a long line of process automation and support technologies. While enterprises that introduce the tool into their operations expect productivity and efficiency gains [19], reports on early adopters temper these expectations and highlight potential or actual risks associated with the technology's implementation [19].

A. Potentially positive aspects of GenAI support on business processes resilience

There are not yet a lot of research findings on the quantitative effects of GenAI on BPR, nevertheless, some reports from business users hint towards positive potentials [20]. Potentially positive impacts of introducing GenAI on BPR include increased and improved output for novices, faster creation, idea generation, and iteration—enhancing *speed* [20] [21]. It potentially boosts *resourcefulness* by providing more variants, alternative patterns, and ideas. Additionally, it offers *flexibility* through alternative generation and aids recovery with templates. *Absorption and rapidity* can be supported by the fast generation of walkaround solutions; the speed of suggestions can also be useful for generating alternative workflows or scenarios in response to a decision, supporting *adaptability* [22] and again resourcefulness. While it can suggest alternatives or leverage scalable processing capacity, GenAI does not add physical or organizational backup capacity, providing limited support for process *redundancy*. Given the indicator definitions, GenAI can provide a focused analysis and, hence, support process *robustness* [23]. Furthermore, updated thresholds and decisions, together with human approval and regular health checks, need to be integrated into the contingency planning. The fast generation speed can

support business impact analysis, and recovery planning to support process *recovery*.

The positive impact of GenAI is, hence, concentrated on content generation to prevent process collapse or restore the process structure based on the original process.

B. Qualitative analysis: Threats to BPR

The introduction of GenAI increases demand for *absorption* within business processes. Employees must handle additional workload from prompt engineering, validation and troubleshooting. Also known as "prompt sprawl" [24] in addition to hallucinations and biased or incorrect outputs. This can lead to technostress [25], AI fatigue [26], and cognitive overload [27], also referred to as "brain fry" [28] [29], thereby reducing individual well-being and, in turn, lowering productivity. Workslop, AI-generated content that lacks substance [30], also endangers productivity and adds to resource consumption. Instead of stabilizing disturbances, processes become strained by the added coordination and control effort [21] [25].

Adaptability may decline when processes struggle to adjust to persistent output inconsistencies due to model drift, i.e., incorrect outputs due to the discrepancy between training data and real-world data [31], and misalignment between model outputs and real-world requirements. Continuous adjustments in workflows, skills, and governance are required, contributing to the productivity J-curve effect, i.e., initial productivity decline during technology adoption as defined by [32]. Fatigue and uncertainty further reduce the system's capacity to adapt effectively. GenAI can enhance process *agility* by enabling rapid output generation for ongoing process steps. This may improve *responsiveness* and short-term *flexibility*, especially in knowledge-intensive or creative tasks. However, *agility* gains are contingent on effective governance of knowledge, including the management of prompts and their interpretation, the curation of input sources, and the quality control of output. Otherwise, faster integration of low-quality outputs can amplify downstream inefficiencies and rework, offsetting productivity gains. *Redundancy* tends to decrease as GenAI replaces or consolidates human tasks, potentially leading to workforce reductions. While this may appear efficient, it reduces fallback options and increases dependency on a single system or vendor (tooling concentration) [33]. Lower redundancy can weaken resilience and create productivity risks if systems fail or outputs degrade. The *recovery* capacity of a process after GenAI introduction can improve if core process structures remain intact and GenAI is layered onto existing routines. In such cases, organizations can revert to established workflows when disruptions occur. However, recovery depends on governance mechanisms such as quality thresholds or fallback procedures. Without them, recovery is slower and more resource-intensive, negatively affecting productivity.

The *resourcefulness* of a business process varies based on the AI literacy of process actors to leverage the added value of the tool and organizational capabilities [34]. High levels of skill, data governance, and decision frameworks support the effectiveness of GenAI, which can further lead to productivity gains. Hence, skill gaps and unclear responsibilities may increase reliance on trial-and-error heuristics, raising coordination costs and reducing efficiency. *Rapidity*, as one of the process resilience characteristics, generally increases with the ad hoc integration of GenAI into workflows, thereby accelerating task execution and content generation response times [35]. However, faster output generation may shift effort toward validating and correcting the provided content [33]. As a result, end-to-end productivity may decline if quality assurance is not properly embedded. Similarly, *robustness* of a business process that integrates GenAI can be threatened by model drift, hallucinations, biased outputs, and vendor dependence. Without clear governance, quality thresholds, and their monitoring, process reliability can decline. This may lead to inconsistent or erroneous outputs and increased troubleshooting, diverting resources away from value creation and reducing overall productivity.

Hence, across all dimensions of business process resilience, integrating GenAI in a non-re-designed business process can introduce a tension between potential efficiency gains and short-term productivity losses: *Productivity* may decrease due to low trust in outputs, increased validation effort, and declining well-being. "Workslop" and inconsistent quality create rework and coordination overhead. *Transition costs* (skills, governance, process redesign) reinforce the productivity J-curve. Without systemic implementation of data governance, quality assurance roles, and decision frameworks, GenAI shifts employees' effort from value creation to error correction.

In this case, GenAI functions less as a simple productivity enhancer and more as a stress test of business process resilience, because productivity gains depend on whether organizations can stabilize and govern the disruption it introduces. The qualitative analysis across BPR characteristics emphasizes the process's ability to absorb and adapt to the disruption caused by the introduction of GenAI tools. Rather than treating GenAI as a purely technological upgrade, it frames the change as a stress test of process resilience, revealing whether the process can remain stable as it adjusts to new forms of work.

To illustrate the impact of incorporating GenAI into a business process in relation to business performance indicators, a hypothetical scenario is described and simulated in the following section. Focusing on process times, costs, and resources enables an assessment of the technology's impact on process performance and structure.

IV. Example Assessment of a Business Process from Marketing Domain

In the following, a hypothetical scenario in the context of a medium-sized enterprise, involving the introduction of GenAI for typical business tasks, is described. The impact of the tool's introduction on business process resilience (BPR) is analyzed in light of the BPR characteristics described above. For a better overview, the process is modeled in BPMN before (Figure 1) and after (Figure 2) the GenAI

integration. Subsequently, a simulation is run for both variants, and performance indicators are compared.

A. Qualitative Analysis of Business Process Resilience

A medium-sized enterprise that designs sports shoe wear is introducing a GenAI tool to support the marketing department in designing and monitoring a launch campaign for a new product. GenAI is to be used here to generate text and images for the campaign, as well as to support its planning by suggesting the rollout process and performance indicators. The rationale for introducing the technology is to help compose text and image content more quickly and to support project management by structuring the project and its indicators. GenAI, being a new tool, needs to be integrated into the current process. Hence, to create the envisioned output with at least the same quality as before, the process needs to *absorb* the tool into its tasks. It will take time for employees to master the new software tool and the interaction, as well as to *adapt* its output to the original process.

Current reports on industrial use cases describe the use of GenAI for brainstorming ideas and generating marketing content, with improved content *quality* and *time* saving [35] [36]. On the other hand, they also outline a rising workload created by unsatisfactory GenAI output, also referred to as workslop [30] as well as reduced worker wellbeing [21] and increased fatigue [26]. These circumstances can exacerbate their negative effects on *productivity* [32] and workers' *trust* [37]. Alternatively, having potentially an additional resource that can generate required output according to even short-term requirement changes allows the enterprise to rapidly respond to the changes, if the tool and its output are adequately integrated into the process, i.e., quality is assured, and guardrails against false or incorrect content are established. The tool might also enable the implementation of different process variants in response to, e.g., changes in budget and target group, while partial resource *redundancy* may necessitate layoffs in the marketing department. Using GenAI for campaign planning may enable broader risk consideration and thus facilitate problem diagnosis in advance. *Rapidity* in response can be supported by faster content generation or adjustments to planning steps within the campaign rollout. From a data-processing perspective, to avoid data and *model drift* after introducing guardrails against wrong, off-brand, or non-compliant outputs, content creation should be supported by curated data and knowledge sources, with regular audits or quality assurance checks in place. It implies involving human expertise to review and compare the output against the expected thresholds.

On the level of human resources and quality, management and maintenance of created knowledge, such as created prompts, approved metadata, and domain- and problem-specific documents, needs to be established to maintain the quality level required for the created content. Automated checks and crisis prompts can help maintain core *quality* under pressure. Model drift, outrages, and policy shifts need to be managed through contingency planning. Maintaining the data sources and prompt libraries requires resources and effort that drain on the *redundancy*, *resourcefulness*, and *flexibility*.

To complement the BPR analysis, a scenario involving the marketing process of marketing campaign design and launch is presented in the following section. Exploring the scenario in terms of process performance metrics and business process resilience can inform managerial decisions about introducing GenAI into a working process.

B. Process simulation

The marketing campaign planning process was modeled both without (Figure 1) and with (Figure 2) GenAI integration. Figure 2 reveals an elevated number of decisions, rework loops, and lane changes in the GenAI-supported process. The model was enriched with process times to enable the process simulation and calculation of performance indicators such as time, costs, and resource utilization.

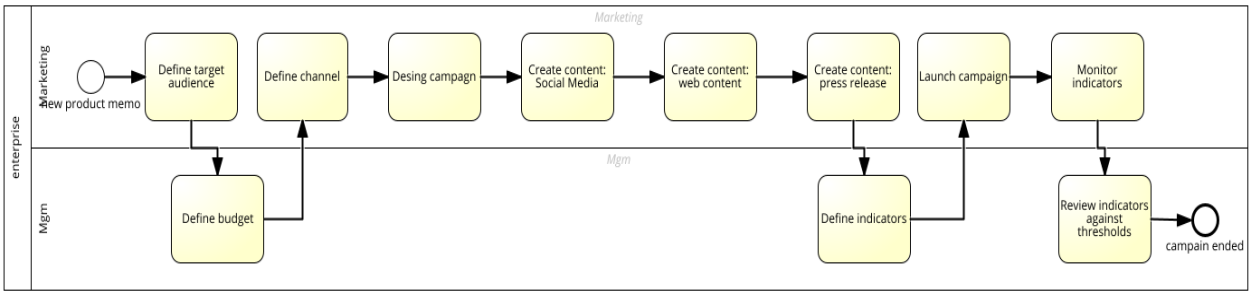


Figure 1. BPMN model of the original process.

The simulation was run using Signavio, a business process modeling tool that also provides process simulation functions. In the process model without GenAI-supported content generation, the duration of tasks for human actors was set to 2 hours, and the duration of decision tasks was set to 1 hour. In the process model with GenAI support, the duration of the generation tasks performed by the GenAI tool was set to 10 minutes, based on the assumption that multiple

prompts were needed to produce a feasible output and reported time-saving rates [36]. For rework loops, the GenAI-created content was set to a 60% rework probability to allow for quality assessment and prompt adjustment. In both scenarios, the marketing employee's rate was set at 50€ per hour, and the manager's at 90€ per hour. No costs were considered for the GenAI tool, as the amount of tokens used for the campaign was considered as negligible.

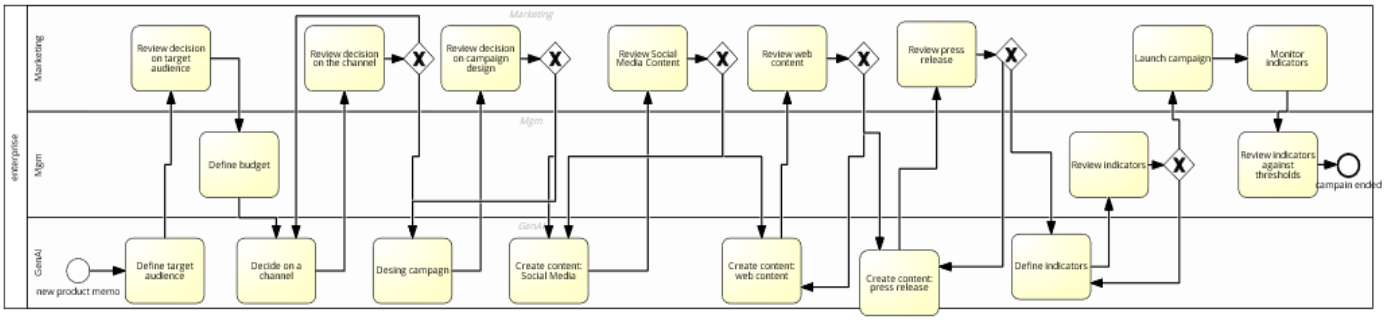


Figure 2. BPMN model of the GenAI-supported process.

While the performance metrics in the simulation are estimates, considering their relations provides insights that warrant further research. The *processing time* was calculated as the average across five campaigns because the tool's five-day workweek setting was used. Process management and marketing employees were identified as *bottlenecks* in the original process through simulation. The average processing time of the original process was calculated as 2 days, 14 hours, and 45 minutes. Here, the average *resource utilization* was 97% for marketing employees and 59% for management employees (see Table 1).

In the Gen-AI-supported scenario, only the marketing employee is identified as the *bottleneck*, and the processing time is calculated to be 3 days, 3 hours, and 40 minutes. *Resource utilization* is at 96% for the marketing employee, 63% for the management employee, and 34% for the GenAI tool.

TABLE I. SIMULATION RESULTS

Performance indictor	***Original process***	***GenAI-supported process***
Processing time	2d:14h:45minutes	3d:3h:40 minutes
Bottleneck	Marketing, Management	Marketing
Costs	4.088	4.125€
Resource utilization		
Marketing	97%	96%
Management	59%	63%
GenAI	NaN	34%

The simulation of the hypothetical business process pre- and post-introduction of the GenAI tool indicates that more resources are utilized, without a reduction in the resource utilization rate for human actors, when GenAI tools are involved. The simulation revealed a slight rise in the utilization rate for the management employee. Workslop and the increase in the resource utilization rate can be attributed to the reworking loops, indicating the impact on *absorption* and *flexibility* characteristics of BPR. The increased processing time after GenAI integration, while not representative, raises the question of GenAI's added value with respect to the characteristic of *rapidity*. The addition of the tool has positive implications for *redundancy*, but the resource utilization rate indicates that both human involvement and tool support are needed to deliver process results, thereby requiring more resources.

Performing the process simulations has indicated a potential discordance: Introducing GenAI into the existing process without changing knowledge management, data governance, or process logic structure affects BPR. It also has a potentially negative effect on the human resources involved in the process, undermining the positive effects of BPR and business performance in terms of time, costs, and resource utilization.

V. DISCUSSION

This research examines how integrating GenAI into existing business processes affects Business Process Resilience (BPR). It views this as a challenge that tests BPR's robustness rather than its productivity enhancement. GenAI is seen as a disruptive intrusion, and its benefits depend on an organization’s ability to adapt through good governance, clear usage policies, clear task definition, and strong knowledge management. The preliminary process analysis suggests that traditional process performance metrics, such as time, costs, and resources, are insufficient for evaluating post-GenAI process performance; measuring process and output quality, as well as monitoring human involvement to balance the negative effects on process quality, is necessary to identify trade-offs. While GenAI seems to boost ideation and support, it offers limited redundancy and might cause rework, increasing resource use, especially with weak governance. Furthermore, process simulation shows no clear cost or time improvements and

may increase process dependency and quality risks, such as bias and wrong content. Further factors that can diminish productivity include increased workload and technostress, with human roles shifting from creation to oversight, requiring adaptation of their skills, tasks, and responsibilities.

VI. Conclusion and Future Work

In this paper, the question of the potential impact of GenAI introduction on BPR in a non-re-designed process was examined using a qualitative approach of content mapping and process simulation. While the described results and insights stem from a qualitative analysis of current industry reports on GenAI effects on productivity and workforce, and the process simulation is based on rough quantitative estimates from the industry, the approach nevertheless illustrates how combining quantitative and qualitative insights can guide the assessment of technology implementation into the existing process. It became visible that business process resilience in the analyzed context depends on safeguards such as quality assurance, curated knowledge, process re-design, and re-skilling. Overall, the findings suggest that GenAI impacts BPR by shifting process resources towards governance, quality assurance, and oversight rather than mere automation. Hence, future work will aim to operationalize BPR characteristics and gather empirical data on GenAI's impact on process performance and resilience.

References

[1] A. Kraus, J.-R. Rehse, and H. van der Aa, "Data-driven assessment of business process resilience," Process Sci, vol. 1, no. 1, p. 4, Oct. 2024, doi: 10.1007/s44311-024-00004-2.

[2] G. Müller, T. G. Koslowski, and R. Accorsi, "Resilience - A New Research Field in Business Information Systems?," in *Business Information Systems Workshops*, W. Abramowicz, Ed., Berlin, Heidelberg: Springer, 2013, pp. 3–14. doi: 10.1007/978-3-642-41687-3_2.

[3] K. Thoma, B. Scharte, D. Hiller, and T. Leismann, "Resilience Engineering as Part of Security Research: Definitions, Concepts and Science Approaches," Eur J Secur Res, vol. 1, no. 1, pp. 3–19, Apr. 2016, doi: 10.1007/s41125-016-0002-4.

[4] K. Furuta, "Resilience Engineering," in Reflections on the Fukushima Daiichi Nuclear Accident: Toward Social-Scientific Literacy and Engineering Resilience, J. Ahn, C. Carson, M. Jensen, K. Juraku, S. Nagasaki, and S. Tanaka, Eds., Cham: Springer International Publishing, 2015, pp. 435–454. doi: 10.1007/978-3-319-12090-4_24.

[5] M. K. Linnenluecke, "Resilience in Business and Management Research: A Review of Influential Publications and a Research Agenda," International Journal of Management Reviews, vol. 19, no. 1, pp. 4–30, 2017, doi: 10.1111/ijmr.12076.

[6] Y. Sheffi and J. Rice, "A Supply Chain View of the Resilient Enterprise," MIT SMR, Oct. 2005, Accessed: May 08, 2026. [Online]. Available: https://sloanreview.mit.edu/article/a-supply-chain-view-of-the-resilient-enterprise/

[7] A. Challapally, C. Pease, R. Raskar, and P. Chari, "State of AI in Business 2025", MIT NANDA, July 2025.

[8] L. Y. Koh, S. E. Toh, and K. F. Yuen, "The influence of digital drivers on organisational digital resilience: An organisational information processing theory perspective," Technology in Society, vol. 86, p. 103283, Jun. 2026, doi: 10.1016/j.techsoc.2026.103283.

[9] Y. Y. Haimes, "On the Definition of Resilience in Systems," Risk Analysis, vol. 29, no. 4, pp. 498–501, 2009, doi: 10.1111/j.1539-6924.2009.01216.x.

[10] J. Bughin, "Robustness and Renewal as key dynamic capabilities for corporate resilience," European Research on Management and Business Economics, vol. 32, no. 2, p. 100308, May 2026, doi: 10.1016/j.iedeen.2026.100308.

[11] S. Duchek, "Organizational resilience: a capability-based conceptualization," Bus Res, vol. 13, no. 1, pp. 215–246, Apr. 2020, doi: 10.1007/s40685-019-0085-7.

[12] R. M. Zahoransky, C. Brenig, and T. Koslowski, "Towards a Process-Centered Resilience Framework," in 2015 10th International Conference on Availability, Reliability and Security, Aug. 2015, pp. 266–273. doi: 10.1109/ARES.2015.68.

[13] M. Bhuiyan, M. M. Z. Islam, G. Koliadis, A. Krishna, and A. Ghose, "Managing Business Process Risk Using Rich Organizational Models," in 31st Annual International Computer Software and Applications Conference - Vol. 2 - (COMPSAC 2007), Beijing, China: IEEE, Jul. 2007, pp. 509–520. doi: 10.1109/COMPSAC.2007.138.

[14] J. Lee, S. Lee, J. Kim, and I. Choi, "Dynamic human resource selection for business process exceptions," Knowl Process Manag, vol. 26, no. 1, pp. 23–31, Jan. 2019, doi: 10.1002/kpm.1591.

[15] J. W. Kule, "Absorptive Capacity as a Catalyst for Resilience in SMEs: A Literature Analysis," JRIIE, Apr. 2025, doi: 10.59765/mgpr6382.

[16] K. Meechang, K. Medini, and M. Pero, "Measuring Agility and Resilience in Engineer-to-Order Contexts," in Advances in Production Management Systems. Cyber-Physical-Human Production Systems: Human-AI Collaboration and Beyond, H. Mizuyama, E. Morinaga, T. Nonaka, T. Kaihara, G. von Cieminski, and D. Romero, Eds., Kamakura: Springer Nature Switzerland, 2025, pp. 71–84. doi: 10.1007/978-3-032-03542-4_5.

[17] R. Cognini, F. Corradini, S. Gnesi, A. Polini, and B. Re, "Research challenges in business process adaptability," in Proceedings of the 29th Annual ACM Symposium on Applied Computing, in SAC '14. New York, NY, USA: Association for Computing Machinery, Mar. 2014, pp. 1049–1054. doi: 10.1145/2554850.2555055.

[18] A. Sarkar, "Will Code Remain a Relevant User Interface for End-User Programming with Generative AI Models?," in Proceedings of the 2023 ACM SIGPLAN International Symposium on New Ideas, New Paradigms, and Reflections on Programming and Software, Cascais Portugal: ACM, Oct. 2023, pp. 153–167. doi: 10.1145/3622758.3622882.

[19] M. Nguyen, N. Trinh, A. Mehrotra, and S. Basahel, "Generative AI in the workplace: how employee experiences influence work outcomes?," Journal of Enterprise Information Management, vol. 38, no. 5, pp. 1647–1666, May 2025, doi: 10.1108/JEIM-11-2024-0637.

[20] M. Mas-Machuca, A. Akhmedova, and F. Marimon, "Generati/ve AI and Workplace Productivity: A Qualitative Study in Spain," pp. 625–631, 2025.

[21] J. Jia, X. Ning, and W. Liu, "The consequences and theoretical explanation of workplace AI on employees: a systematic literature review," J. Digit. Manag., vol. 1, no. 1, p. 14, Oct. 2025, doi: 10.1007/s44362-025-00016-3.

[22] S. Gao, P.-L. Teh, and H. H. P. Ho, "Digital transformation and innovation in small and medium enterprises (SMEs): a systematic review and future research agenda," Cogent

Business & Management, vol. 13, no. 1, p. 2612775, Jan. 2026, doi: 10.1080/23311975.2026.2612775.

[23] "5 Generative AI Resilience Use Cases Transforming Business Continuity," Disaster Recovery Journal. Accessed: May 01, 2026. [Online]. Available: https://drj.com/journal_main/generative-ai-resilience-use-cases/

[24] "Prompt Sprawl: What the Real Costs Look Like in Production," DEV Community. Accessed: Apr. 27, 2026. [Online]. Available: https://dev.to/gorealai/prompt-sprawl-what-the-real-costs-look-like-in-production-3mo9

[25] S. Zhang, P. Guo, Y. Yuan, and Y. Ji, "Anxiety or engaged? Research on the impact of technostress on employees' innovative behavior in the era of artificial intelligence," Acta Psychologica, vol. 259, p. 105442, Sep. 2025, doi: 10.1016/j.actpsy.2025.105442.

[26] M. Ragolane and S. Patel, "Too Much, Too Fast: Understanding Ai Fatigue In The Digital Acceleration Era," International Journal of Arts Humanities & Social Science, vol. 6, pp. 53–60, Aug. 2025, doi: 10.56734/ijahss.v6n8a7.

[27] "AI and the Rise of Cognitive Overload | College of Public Health." Accessed: Apr. 13, 2026. [Online]. Available: https://publichealth.gmu.edu/news/2026-03/ai-and-rise-cognitive-overload

[28] J. Bedard, M. Kropp, M. Hsu, O. T. Karaman, J. Hawes, and G. R. Kellerman, "When Using AI Leads to 'Brain Fry,'" Harvard Business Review, Mar. 05, 2026. Accessed: Apr. 13, 2026. [Online]. Available: https://hbr.org/2026/03/when-using-ai-leads-to-brain-fry

[29] A. Wray and P. Merton, "'Brain fry' in Just a Minute : the challenges of talking without hesitation, repetition or deviation," Comedy Studies, vol. 15, no. 2, pp. 137–153, Jul. 2024, doi: 10.1080/2040610X.2024.2373579.

[30] K. Niederhoffer, G. R. Kellerman, A. Lee, A. Liebscher, K. Rapuano, and J. T. Hancock, "AI-Generated 'Workslop' Is Destroying Productivity," Harvard Business Review, Sep. 22, 2025. Accessed: Feb. 23, 2026. [Online]. Available: https://hbr.org/2025/09/ai-generated-workslop-is-destroying-productivity

[31] "Understanding Model Drift and Data Drift in LLMs (2026 Guide)." Accessed: May 05, 2026. [Online]. Available: https://orq.ai/blog/model-vs-data-drift

[32] E. Brynjolfsson, D. Rock, and C. Syverson, "The Productivity J-Curve: How Intangibles Complement General Purpose Technologies," American Economic Journal: Macroeconomics, vol. 13, no. 1, pp. 333–372, Jan. 2021, doi: 10.1257/mac.20180386.

[33] J.-E. D. Neve, J. T. Hancock, and K. Niederhoffer, "Why Companies That Choose AI Augmentation Over Automation May Win in the Long Run," Harvard Business Review, Apr. 15, 2026. Accessed: Apr. 23, 2026. [Online]. Available: https://hbr.org/2026/04/why-companies-that-choose-ai-augmentation-over-automation-may-win-in-the-long-run

[34] M. Hoffmann, S. Boysel, F. Nagle, S. Peng, and K. Xu, "Generative AI and the Nature of Work," HBS Working Paper Series, no. 25–021, 2025.

[35] L. Ma, P. Yu, X. Zhang, G. Wang, and F. Hao, "How AI use in organizations contributes to employee competitive advantage: The moderating role of perceived organization support," Technological Forecasting and Social Change, vol. 209, p. 123801, Dec. 2024, doi: 10.1016/j.techfore.2024.123801.

[36] N. Rudan, "6 Ways Marketers Are Using Generative AI: Is It Really Saving Time?," Databox. Accessed: May 02, 2026. [Online]. Available: https://databox.com/how-are-marketers-using-gen-ai

[37] WEF, "World Economic Forum," White paper, 2023. Accessed: Apr. 19, 2026. [Online]. Available: https://www3.weforum.org/docs/WEF_Measuring_Digital_Trust_2023.pdf

Inferring Political Orientation from Credit Score-Relevant Variables

An Empirical Study on Profiling and Proxy Inference Through Sensitive Attributes

David Schnepp
Department of Business and Management
Brandenburg University of Applied Science
Brandenburg a. d. Havel, Germany
Email: schnepp@th-brandenburg.de

Olga Levina
Department of Business and Management
Brandenburg University of Applied Science
Brandenburg a. d. Havel, Germany
Email: levina@th-brandenburg.de

Abstract— **The increasing availability of data in everyday life expands the possibilities for profiling individuals across social, economic, and political domains. Even when sensitive attributes are not explicitly collected, they may be inferred from seemingly non-sensitive demographic and socioeconomic information. This paper explores whether political orientation can be predicted from credit-related variables, motivated by concerns that algorithmic systems might infer sensitive attributes from seemingly non-sensitive data. Using data from the European Social Survey (Round 10), voting behavior in the 2021 German federal election is treated as a multiclass classification problem. Several common supervised learning methods, such as logistic regression, support vector machines, k-nearest neighbors, and boosting-based decision trees, are employed. The results indicate that political orientation cannot be reliably predicted in this context. Nonetheless, this does not eliminate the broader risk of proxy-based inference. Moreover, inferability depends not only on the variables themselves, but also on the broader analytical context. The literature review indicates that differences in data availability, feature construction, preprocessing, and model training approaches can significantly influence information gain and predictive inferability. Instead, the findings emphasize the importance of systematic, context-specific assessments of inferability, with significant consequences for data protection and AI regulation.**

Keywords - Machine Learning; Profiling; Inference of sensitive attributes; Proxy discrimination.

I. INTRODUCTION

The increasing deployment of automated, data-driven systems across domains such as finance, public administration, and political processes fundamentally reshapes how personal information is processed and interpreted [1][2]. It also challenges the assumption that data used in such systems can be clearly categorized as sensitive or non-sensitive [3]. Data is not neutral [4]. Additional information can be derived from aggregating, combining, and analyzing the collected data, often beyond the original purpose of data collection [1][5].

This raises a central societal concern. Sensitive personal attributes do not need to be explicitly collected to be effectively used. Yeom, Datta, and Fredrikson, for example, demonstrated that combinations of ostensibly non-sensitive variables in predictive policing-related datasets can function as powerful proxies for racial composition and reinforce discriminatory outcomes in algorithmic decision-making. In their analysis, a proxy constructed from 58 seemingly unrelated features exhibited a particularly strong association with race [6]. They can be reconstructed as so-called proxy variables from data that were originally considered non-sensitive, giving rise to new forms of indirect discrimination. They enable the inference of characteristics such as political orientation, ethnicity, or religion. As a result, individuals may be subject to differential treatment based on attributes that were neither disclosed nor directly processed [7].

The protection concern thus shifts from the question of which data is collected to what information can be inferred from it. This shift has profound implications for data protection and fairness, as it undermines traditional regulatory approaches that focus primarily on input data categories. Algorithmic inference of sensitive attributes therefore represents not only a technical challenge but a structural risk for discrimination in increasingly data-driven societies [5] [7][8].

The research question addressed in this study is thus whether the political orientation of individuals, a notably safeguarded sensitive characteristic, can be inferred from non-sensitive, credit-relevant attributes using standard supervised learning models.

Finding an affirmative response would suggest that based on non-sensitive, credit-relevant data an AI system can potentially leverage individual political preferences. Regulatory frameworks such as the EU AI Act (AIA) and the General Data Protection Regulation (GDPR) seek to establish safeguards for sensitive personal data, particularly by restricting their collection and use in automated decision-making contexts [9][10]. However, these protections become less straightforward when sensitive attributes are not directly processed but can be inferred from non-sensitive data. The European Data Protection Board addresses this challenge by clarifying that profiling, defined as the automated processing of personal data to evaluate or predict individual characteristics, also encompasses situations in which sensitive information is inferred. While this interpretation broadens the scope of data protection law, it still raises practical and conceptual challenges regarding the effective regulation and enforcement of such inference-based processing [5].

To address this question, an empirical investigation was conducted using the European Social Survey (ESS) Round 10. The target variable is voting behavior in the 2021 German

federal election, which results in a multiclass classification problem [11]. Core supervised-learning algorithms were used for processing the non-sensitive attributed: logistic regression, Support Vector Machines (SVM), k-Nearest Neighbors (KNN), and a boosting-based decision tree.

The empirical results consistently show across all model types that no reliable prediction of political orientation is possible. Nevertheless, this analysis demonstrates that systematic approaches to estimate ex ante the likelihood of attribute inference under different conditions are needed in research and practice. The inability to infer political orientation in this setting does not guarantee safety in others; practitioners should not interpret negative results as proof of absence of profiling risk and regulators may need to consider implementing flexible, case-by-case evaluation mechanisms instead of rigid assumptions about inferability.

The contribution of this research is twofold. First, an empirical assessment shows that, in this dataset and under this label definition, predictive performance remains close to simple baselines, suggesting limited inferability from credit-relevant attributes alone. Second, the scope of this finding is clarified: it does not generalize to all contexts. Inferability is shaped by the feature space, label granularity, sample size, class balance, and modeling choices. We therefore recommend systematic, context-specific inferability audits as a prerequisite for claims about the feasibility or safety of inferring sensitive attributes. This targeted evidence can help calibrate regulatory discussions that currently alternate between universal warnings and untested assurances.

The remainder of the paper is structured as follows: First, related work on the inference of sensitive attributes is reviewed. This is followed by a description of the study's methodology. Subsequently, the empirical results are presented and discussed, considering their implications.

II. Data Processing and Sensitive Attributes: State of the Art

Empirical research provides substantial evidence that sensitive personal attributes can be inferred from non-sensitive data with considerable accuracy. A prominent example is the work of Kosinski et al. [12], which demonstrates that digital behavioral traces such as Facebook Likes can predict a wide range of sensitive attributes. Exploiting a dataset of approximately 58,000 individuals (170 likes on average) and applying logistic regression for dichotomous variables and linear regression for numeric variables combined with dimensionality reduction, the study achieved high predictive performance across multiple categories. Both regression models used 10-fold cross-validation and k=100 SVD components. Area Under Curve (AUC) values reached 0.95 for distinguishing between Caucasian and African American individuals, 0.82 for religion (Christianity vs. Islam), and 0.85 for political orientation (Democrat vs. Republican). These results highlight the strong inferential power of seemingly non-sensitive behavioral data. Performance for numeric features was also partly strong. Age was predicted with a Pearson correlation of 0.75, whereas predictions for the number of Facebook friends were less accurate, with a Pearson correlation of 0.47.

Similar findings emerge in financial contexts. In a study by Hassani [13], ethnicity was predicted from standard credit-scoring variables such as income, credit limits, and account balances. The dataset comprised about 400 cases covering African-American (99), Asian (102), and Caucasian (199) individuals. Applying a Random Forest model (750 trees) and a 75/25 train-test split, the study achieved an F1-score of approximately 0.65 on the original dataset, which increased to around 0.70 under modified conditions and up to 0.99 after applying the Synthetic Minority Oversampling Technique (SMOTE) to rebalance the dataset by creating synthetic data. The results indicate that financial variables can act as strong proxies for ethnicity, although the study is limited by its small sample size and reliance on data manipulation techniques.

More recent work further demonstrates the inferability of sensitive attributes across data modalities. Chaturvedi & Chaturvedi [14] show that religion can be predicted with very high accuracy from personal names alone. Employing a large-scale dataset of over 115,000 households from India and an 80-10-10 train-validation-test split, they applied several machine learning models, including dictionary-based approaches, language models, logistic regression, SVM, and Convolutional Neural Network (CNN) sequencing models. Evaluation was performed with F1, precision, and recall. The study achieved strong predictive performance, exceeding that of typical dictionary-based approaches. The best predictive performance for single names was obtained by CNN with F1 of 95.86, while for concatenated names (containing more information) SVM outperformed CNN with F1 of 97.33. The results illustrate that even basic identifiers contain rich latent information that can be exploited for sensitive inference.

Beyond explicitly sensitive targets, related work also demonstrates the predictive power of behavioral financial data for complex personal characteristics. Kim et al. [15] applied large-scale open-banking transaction data comprising approximately 100,000 individuals and 180 million transactions to model financial vulnerability. Combining open-banking data with typical credit-scoring approaches opens wide potential for credit risk valuation. Applying machine learning models such as Random Forest, logistic regression, and XGBoost, the study achieves an accuracy of around 0.77 and an F1-score of approximately 0.62 in identifying people with disabilities. They used an 80/20 train-test split with hyperparameter tuning via grid search and 10-fold cross-validation.

The aforementioned studies underscore that the efficacy of sensitive attribute reconstruction is fundamentally contingent on the alignment between the model architecture and the underlying data structure. This architecture-performance gap is particularly evident in Chaturvedi & Chaturvedi [14], where CNNs excelled at extracting features from sequential name patterns (F1: 95.86), whereas SVMs demonstrated superior performance on high-dimensional, concatenated datasets (F1: 97.33). Such variations indicate that additional data only enhances predictive performance if the model has sufficient complexity to process the increased information density. The application of resampling strategies, such as SMOTE in

Hassani [13], demonstrates that addressing class imbalances can drastically improve model performance, increasing F1 Scores from 0.65 to 0.99 by enabling algorithms to learn patterns in minority classes that would otherwise be obscured by statistical noise. The inferential power of these models is highly domain-specific and depends on the quality of proxy variables. The selective predictive power observed in Kosinski et al. [12] illustrates this limitation. While Facebook Likes serve as a high-performance proxy for ethnicity (0.95 AUC), they are less informative for quantitative social metrics such as friend count (0.4 correlation).

The existing research suggests that the transition from non-sensitive input to sensitive inference is governed by a rigorous methodological framework involving individual feature pre-processing (i.e., dimensionality reduction), cross-validation, and systematic hyperparameter tuning. However, the entire prediction process is highly individualized, based on the dataset and (target) sensitive variables, which collectively challenge current technical and regulatory safeguards [12][13][14][15].

III. RESEARCH DESIGN

The research question of this study is whether individuals' political orientation, operationalized via voting behavior as a particularly sensitive attribute, can be reconstructed using established supervised learning models based solely on legitimate, credit-relevant features. These models were selected as established approaches in applied machine learning for structured data, without relying on highly complex model architectures.

To address this question, an empirical investigation was conducted using the European Social Survey (ESS) Round 10. The ESS is a social science dataset that includes demographic and socioeconomic information, as well as respondents' political orientations, such as their party affiliation and views on specific political issues, e.g., the government's role in reducing income inequality [16]. For the analysis, 15 features relevant to real creditworthiness assessments were selected from the dataset, as shown in Table I.

TABLE I. OVERVIEW OF SELECTED FETAURES AND DATA TYPES.

ESS-Variable	Credit-Scoring related feature	Data Type
prtvfde2	Voting behavior (target var)	Nominal
gndr	Gender	Nominal
agea	Age	Metric
domicil	Domicile / housing location	Ordinal
rshpsts	Relationship status	Nominal
hhmmb	Household size	Metric
hhcd	Children living in household	Metric
wrkctra	Contract type	Nominal
educde1	General education level	Ordinal
edubde2	Vocational education level	Ordinal
hinctnta	Household income	Ordinal
hincsrca	Main source of income	Nominal
nacer2	Industry sector	Nominal
isco08	Occupation	Nominal
ctzcntr	Citizenship	Nominal
brncntr	Country of birth	Nominal

The selection of input features followed a structured three-step approach. First, relevant variables were identified based on established criteria from the literature on creditworthiness assessments and complemented by practical data requirements derived from two real-world credit application forms. These criteria were then systematically mapped onto the ESS dataset by identifying where possible and valid approximations. This applies, for instance, to credit scoring categories such as address or the number of children in a household. The *address* category was mapped to *domicil*, whereas the number of children living in the household was determined by aggregating the variables for *persons in household* in combination with the *relationship to respondent* indicators (*rshipa2–rshipa15)* [16]. Voting behavior in the 2021 German federal election was used as the target variable, resulting in a multiclass classification problem. The ESS variable prtvfde2 was restricted to seven valid party classes: *1 = CDU/CSU (15.9%), 2 = SPD (20.2%), 3 = Die Linke (3.9%), 4 = Bündnis 90/Die Grünen (15.4%), 5 = FDP (9.8%), 6 = AfD (3.8%), and 7 = Andere (4.9%).* Responses without substantive informational content, including 66 = not applicable, *77 = refusal, 88 = don't know,* and *99 = no answer*, were recoded as NaN and excluded from the modeling dataset. Since supervised classification requires a valid target label, only observations with values between 1 and 7 were retained for model training.

The analytical approach is based on the Cross-Industry Standard Process for Data Mining framework, a process model for structuring machine learning projects [17]. All data processing were implemented in Python within Jupyter Notebook environments using pandas for data manipulation and scikit-learn for machine learning workflows. Four supervised learning algorithms were used: logistic regression, Support Vector Machines (SVMs), k-Nearest Neighbors (KNN), and a boosting-based decision tree (HistGradientBoostingClassifier). Data pre-processing included the encoding of categorical and ordinal variables as well as the handling of missing values through variable-specific imputation strategies. For logistic regression, SVM, and KNN, categorical variables were transformed using one-hot encoding to obtain a numerical representation suitable for model training. Numerical predictors were subsequently scaled according to model-specific requirements. StandardScaler was applied for logistic regression and SVM, whereas MinMaxScaler was used for KNN due to its distance-based nature. For the boosting-based decision tree model, categorical variables were retained as categorical data types where applicable, allowing them to be processed without one-hot encoding. All preprocessing operations were integrated directly into the machine learning pipelines to ensure a standardized structure of the input matrices and to prevent data leakage during cross-validation. Particular attention was given to the stratified train (0.8)-test (0.2) split to enable evaluation on independent data and to control for variance and bias. Due to preprocessing procedures, the final analytical sample used for model training comprised n=6,443 observations.

For hyperparameter optimization, GridSearchCV was used to perform a structured, exhaustive search across the

entire defined parameter space. Systematic hyperparameter variations were evaluated. For SVM, the grid search explored variations in C, gamma, and different kernel types, while for k-NN it evaluated adjustments to n_neighbors, distance metrics, and weighting schemes. Each configuration was tested in experimental setups, allowing the performance of all parameter combinations to be assessed under consistent conditions and enabling the identification of the optimal model configuration. To obtain robust and distributionally faithful estimates of generalization performance, stratified 5-fold cross-validation was used. Preliminary experiments with stratified 10-fold cross-validation did not yield meaningful performance improvements, while substantially increasing computational complexity. The combination of GridSearch and StratifiedKFold ensures that model variants are optimized independently from the final test data. To support reproducibility, additional details on missing-value handling, imputation strategies, hyperparameter grids, the specific algorithm, and the encoding of isco08 and nacer2 variables are available from the authors upon request.

Model performance was assessed using common classification metrics, particularly accuracy, F1 score, precision, sensitivity, and specificity. The no-information rate, defined as the accuracy achieved by always predicting the most frequent class, served as a reference point. This methodological framework enables a nuanced evaluation, especially in the presence of imbalanced class distributions typically associated with electoral choices.

IV. RESULTS

The empirical results consistently show that no reliable prediction of political orientation is possible across all model types. Across all models, accuracy ranged from 0.246 to 0.316 and thus only marginally exceeded the no-information rate of 0.272 (Fig. 1).

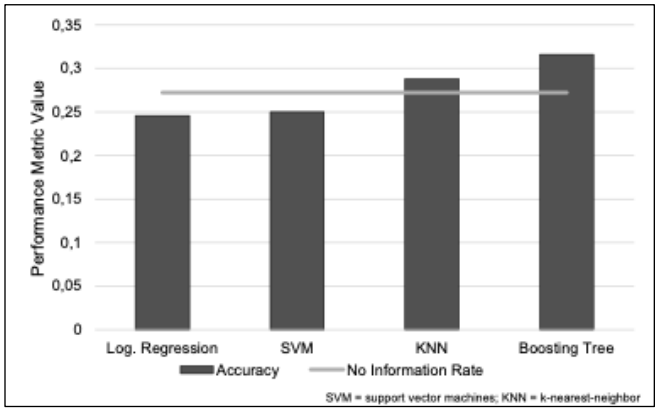


Figure 1. Comparison of Classification Models Accuracy and No Information Rate.

F1-scores, precision, and sensitivity remained consistently low, with particularly poor model performance for smaller parties and rare classes. Regardless of whether the F1-score is computed as a macro or weighted average, the results follow a similar pattern. Macro F1-scores ranged from 0.194 to 0.225, while weighted F1-scores ranged from 0.251 to 0.280, indicating overall weak classification performance (Fig. 2).

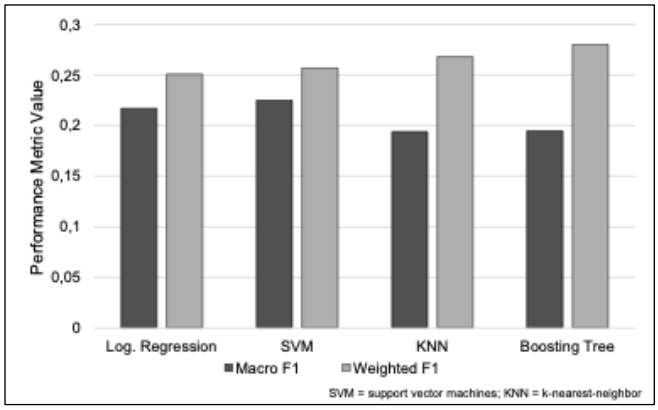


Figure 2. Comparison of Classification Models by F1-Score Metrics.

Macro sensitivity between 0.197 and 0.253 shows that only about one-fifth to one-quarter of actual class instances were correctly identified, and macro precision between 0.222 and 0.234 indicates that fewer than one-quarter of positive predictions were accurate. The models consistently achieved high macro specificity values ranging from 0.870 to 0.874, indicating reliable identification of non-memberships. However, they nonetheless failed to produce meaningful differentiation for positive class assignments (Fig. 3).

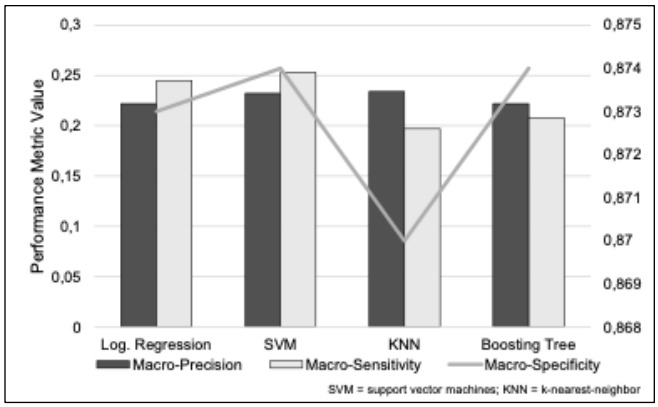


Figure 3. Comparison of Classification Models by Core Performance Metrics.

A permutation-based feature-importance analysis was added to understand the importance of each variable. No feature showed strong predictive relevance; age had the highest relative importance (0.0297 ± 0.0112), followed by general education level (0.0115 ± 0.0081), domicile (0.0109 ± 0.0070), children living in the household (0.0106 ± 0.0030), and household income (0.0099 ± 0.0039). Several variables contributed only marginally or negatively, suggesting that the selected creditworthiness-related features provide limited information for reconstructing political orientation.

V. DISCUSSION

This research addressed the question of whether it is possible to infer voting behavior based on data collected for socio-economic and credit score- related purposes. Popular classification algorithms were used to combine and strucutre the data. The results showed that no reliable prediction of political orientation was possible across the applied model types. The findings of this study therefore indicate that no meaningful inference of political orientation could be drawn from the given setup, suggesting that attribute profiling based on typical features related to creditworthiness is not feasible in this specific context.

However, this result should be interpreted in light of the imbalanced target distribution. Several party classes are sparsely represented, particularly *Die Linke* (3.9%) and *AfD* (3.8%), which may limit the ability of supervised learning models to learn reliable class-specific patterns. Poor performance should therefore not be interpreted solely as evidence of real non-inferability, but also as potentially reflecting methodological constraints resulting from sparse and unevenly distributed target classes. The negative result may therefore be due to several reasons, which have yet to be fully clarified.

Additionally, prior research demonstrates that reconstructing sensitive attributes from non-sensitive data can be highly effective in other settings. This contrast highlights that such inference processes are not universally generalizable in either direction. Rather, their success critically depends on the underlying data, preprocessing, the presence and strength of proxy variables, and the chosen modeling approach [12][13][14][15]. Ultimately, the current findings do not definitively determine whether political orientation can be derived from creditworthiness-related features.

Furthermore, the findings underscore the methodological sensitivity of inference outcomes. Even within the same empirical setting, variations in modeling choices, such as different architectures or resampling strategies, can yield different results. Although this study did not observe meaningful predictive performance, these dependencies indicate that outcomes are not method-invariant. In addition, the reconstructability of sensitive attributes appears to be attribute-specific, as some characteristics may be more easily inferred than others depending on their relationship to available proxy features. Consequently, the absence of predictive performance in a given setup does not imply the absence of attribute profiling risk, because absolute certainty about the absence of attribute profiling risk, or about non-profiling, cannot be established.

This context dependence has important implications for assessing the risks of attribute profiling. The results suggest that the mere availability of non-sensitive data does not automatically enable the reconstruction of sensitive attributes. At the same time, inferability is the mechanism by which non-sensitive variables may become proxies for sensitive attributes, thereby enabling indirect discrimination. However, the variability observed across studies indicates that such risks cannot be reliably excluded ex ante. Consequently, the feasibility of attribute profiling remains inherently difficult to predict, posing a significant challenge for regulatory frameworks such as the GDPR, which aim to govern and restrict such practices.

VI. CONCLUSION AND FUTRE WORK

In this paper, we examined whether creditworthiness-related variables can be used to infer political orientation as a sensitive attribute. Using selected socio-demographic and economic features from the ESS dataset, we trained and evaluated several machine learning models to assess the feasibility of attribute profiling in this specific context. The results indicate that no reliable prediction of political orientation could be achieved across the applied models, suggesting that such inference is not feasible under the given conditions. At the same time, these findings should be interpreted cautiously, as the absence of reliable predictive performance does not necessarily imply the absence of inferability, but may also reflect methodological constraints such as class imbalance.

The results of this study suggest several important directions for future research and regulatory consideration. First, the inherent unpredictability of attribute profiling's feasibility calls for a shift away from binary risk assessments toward more probabilistic, context-sensitive evaluation frameworks. Future work should therefore focus on systematically mapping the conditions under which sensitive attribute inference becomes viable, including the roles of data richness, feature correlations, class distributions, and model complexity. In particular, comparative studies across a wider range of modeling approaches, from classical statistical methods to advanced machine learning architectures such as neural networks, would help to better understand the extent to which inference risks depend on methodological choices. Future studies should also examine resampling strategies such as SMOTE to address class imbalance and assess whether sparse target classes affect inferability outcomes. In addition, datasets with larger sample sizes and potentially richer socioeconomic variables, such as the SOEP dataset provided by the *Deutsches Institut für Wirtschaftsforschung* (DIW), may offer further opportunities to investigate proxy-based inferability in contexts related to creditworthiness assessments.

Second, the observed attribute-specific variability highlights the need for more fine-grained analyses of which categories of sensitive information are most vulnerable to indirect inference. This suggests that regulatory approaches such as the GDPR and AIA may benefit from incorporating differentiated risk assessments rather than treating all sensitive attributes uniformly. Developing standardized benchmarks and evaluation protocols could support more consistent assessments of profiling risks across contexts and studies.

Finally, the findings underscore the importance of adopting precautionary principles in both system design and policy. Given that the absence of evidence for predictive performance does not constitute evidence of absence, future research should explore robust auditing methods and risk mitigation strategies that remain effective under uncertainty. This includes investigating privacy-enhancing technologies,

model-auditing techniques, and data-minimization practices that can reduce the likelihood of unintended attribute inference, even when such risks are not immediately observable.

REFERENCES

[1] X. Gao et al., "Fairness in machine learning: definition, testing, debugging, and application", *Sci. China Inf. Sci.*, vol. 67, no. 9, p. 191201, Sep. 2024, doi: 10.1007/s11432-023-4060-x.

[2] Risk Research, "Der EU AI Act in Kreditinstituten [The EU AI Act in Credit Institutions]", Aug. 2024. Accessed: May 09, 2026. [Online]. Available: https://www.risk-research.de/fileadmin/userdaten/docs/PDF-Dateien/Whitepaper/Risk_Research_Whitepaper_7_EU_AI_Act.pdf.

[3] P. Quinn and G. Malgieri, "The difficulty of defining sensitive data—the concept of sensitive data in the EU data protection framework", Ger. Law J., vol. 22, no. 8, pp. 1583–1612, Dec. 2021, doi: 10.1017/glj.2021.79.

[4] F. Elsafoury, "Diskriminierung | Wenn der Schein trügt – Deepfakes und die politische Realität [Discrimination | When appearances are deceptive – deepfakes and political reality]". Accessed: May 09, 2026. [Online]. Available: https://www.bpb.de/lernen/bewegtbild-und-politische-bildung/556762/diskriminierung/.

[5] Artikel-29-Datenschutzgruppe, "Leitlinien zu automatisierten Entscheidungen im Einzelfall einschließlich Profiling für die Zwecke der Verordnung 2016/679 [Guidelines on automated individual decision-making and profiling for the purposes of Regulation 2016/679]", Oct. 2017. Accessed: May 09, 2026. [Online]. Available: https://dsb.gv.at/sites/site0344/media/downloads/wp251rev01_de.pdf.

[6] S. Yeom, A. Datta, and M. Fredrikson, "Hunting for discriminatory proxies in linear regression models", 2018, arXiv. doi: 10.48550/ARXIV.1810.07155.

[7] A. D. Selbst, D. Boyd, S. A. Friedler, S. Venkatasubramanian, and J. Vertesi, "Fairness and abstraction in sociotechnical systems", in Proceedings of the Conference on Fairness, Accountability, and Transparency, Atlanta GA USA: ACM, Jan. 2019, pp. 59–68, doi: 10.1145/3287560.3287598.

[8] K. Tiwari, N. Sarkar, K. Bisht, and S. Shukla, "A systematic survey on bias and fairness in machine learning", in Data Science and Security, S. Shukla, H. Sayama, K. Tiwari, and J. V. Kureethara, Eds., Lecture Notes in Networks and Systems, vol. 1355. Singapore: Springer Nature Singapore, 2025, pp. 77–91, doi: 10.1007/978-981-96-4883-2_8.

[9] European Data Protection Supervisor, AI Act Regulation (EU) 2024/1689. Luxembourg: Publications Office of the European Union, 2025, doi: 10.2804/4225375.

[10] European Parliament and Council of the European Union, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), May 4, 2016. Accessed: May 09, 2026. [Online]. Available: http://data.europa.eu/eli/reg/2016/679/oj.

[11] European Social Survey European Research Infrastructure (ESS ERIC), "ESS10 - data from Interviewer's questionnaire, edition 3.1". Sikt - Norwegian Agency for Shared Services in Education and Research, 2023, doi: 10.21338/ESS10INTE03_1.

[12] M. Kosinski, D. Stillwell, and T. Graepel, "Private traits and attributes are predictable from digital records of human behavior", Proc. Natl. Acad. Sci., vol. 110, no. 15, pp. 5802–5805, Apr. 2013, doi: 10.1073/pnas.1218772110.

[13] B. K. Hassani, "Societal bias reinforcement through machine learning: a credit scoring perspective", AI Ethics, vol. 1, no. 3, pp. 239–247, Aug. 2021, doi: 10.1007/s43681-020-00026-z.

[14] R. Chaturvedi and S. Chaturvedi, "It's all in the name: a character-based approach to infer religion", Polit. Anal., vol. 32, no. 1, pp. 34–49, Jan. 2024, doi: 10.1017/pan.2023.6.

[15] S. D. Kim, G. Andreeva, and M. Rovatsos, "The double-edged sword of big data and information technology for the disadvantaged: a cautionary tale from open banking", 2023, arXiv, doi: 10.48550/ARXIV.2307.13408.

[16] European Social Survey ERIC (ESS ERIC), "European Social Survey (ESS), Round 10 - 2020", 2022, Sikt - Norwegian Agency for Shared Services in Education and Research, doi: 10.21338/NSD-ESS10-2020.

[17] S. Selle, Data Science Training - Supervised Learning: Ein praktischer Einstieg ins überwachte maschinelle Lernen [A practical introduction to supervised machine learning]. Berlin, Heidelberg: Springer Berlin Heidelberg, 2024, doi: 10.1007/978-3-662-67960-9.