Human-Computer Interaction in the Big Data Era: Systems, Cognition, and Interactive Intelligence

The convergence of Human-Computer Interaction (HCI) and Big Data has introduced new demands for scalability, responsiveness, and cognitive alignment in interactive systems. As data complexity grows and machine learning (ML) models become central to decision-making, user interfaces (UIs) must evolve from static dashboards to dynamic, adaptive environments that support real-time exploration, transparency, and trust. This survey offers an integrative analysis of recent advances at the intersection of HCI and Big Data, focusing on four core dimensions. First, we synthesize scalable visual interaction techniques that address overplotting, high-dimensional embedding, and cross-modal coordination. Next, we examine system architectures that prioritize interaction responsiveness through adaptive pipelines, decentralized execution, and interface-centric data shaping. Also, we explore cognitive modeling strategies for intent inference, cognitive load detection, and adaptive composition of views. Finally, we evaluate mechanisms for explainability and trust, including interactive explanations, selective transparency, and auditable system behaviors. Together, these contributions define a design agenda for future systems that are not only data-intensive but also human-aware, accountable, and ethically aligned.

CCS Concepts: • **Human-centered computing** → **Visualization systems and tools**; • **Computing methodologies** → *Cognitive science*; Machine learning; • **Applied computing** → Computer-assisted instruction; • **Information systems** → Data analytics; *Big data systems.*

ACM Reference Format:

. 2025. Human-Computer Interaction in the Big Data Era: Systems, Cognition, and Interactive Intelligence. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (CHIGREECE '25).* ACM, New York, NY, USA, 13 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

The proliferation of data in both volume and complexity has fundamentally altered the landscape of HCI. As users increasingly engage with vast, heterogeneous, and dynamically evolving datasets, traditional interaction paradigms fail to support the interpretive, exploratory, and decision-making needs of modern data-intensive tasks. At the same time, Big Data infrastructures—designed for scalability, throughput, and automation—often lack mechanisms to accommodate human cognition, adaptive interaction, or interface-level reasoning. The intersection of these domains demands a shift from isolated system optimisation to holistic, user-aware data environments that integrate visual, architectural, cognitive, and ethical dimensions into their core design [23, 73].

This transformation is especially urgent as analytic systems incorporate ML models whose predictions carry operational, clinical, or societal consequences. In such settings, interactivity cannot be reduced to graphical UI (GUI) design or usability heuristics—it must encompass latency resilience, intent inference, and trust calibration, all integrated across distributed architectures. Building effective interfaces in Big Data environments thus requires coordinated progress in data visualisation, systems engineering, cognitive modelling, and algorithmic explainability. However,

Author's Contact Information:

existing research remains fragmented, with few works offering a unified perspective on how HCI can evolve to meet the challenges of Big Data systems [28, 34].

While significant technical progress has been made in visual analytics, distributed systems, and human-centred artificial intelligence (AI), the absence of a cohesive framework connecting HCI principles to Big Data architecture limits both system usability and human interpretability. Current approaches often treat interactivity as an ancillary concern, bolted onto infrastructures optimised for scale rather than engagement. This has led to a proliferation of powerful tools in computation, but has been weak in terms of transparency, adaptability, and cognitive support. Addressing this disconnect requires a rethinking of design priorities—placing the user not at the edge but at the centre of the Big Data pipeline.

This survey offers a structured and integrative overview of current advances at the intersection of HCI and Big Data. It defines a multi-layered framework that connects visual interaction, system architecture, cognitive modelling, and interpretability, highlighting both state-of-the-art methods and open research challenges. Specifically, the present survey:

- Synthesises advances in scalable visual interaction techniques, including multiscale encoding, progressive rendering, and cross-view coordination for high-dimensional data exploration.
- Examines system architectures designed to support real-time interactivity, such as adaptive query pipelines, decentralised execution frameworks, and interface-centric data shaping.
- Analyses models for cognitive adaptation and user intent inference, highlighting how systems modulate complexity and structure based on user behaviour and cognitive load.
- Explores mechanisms for explanation and trust, covering interactive model introspection, selective transparency, and compliance-aware auditability in human-AI systems.

The remainder of this paper is structured as follows. Section 2 presents scalable visual interaction techniques. Section 3 discusses human-centered system architectures. Section 4 explores cognitive adaptation and user modeling. Section 5 examines explainability, trust, and auditability. Section 6 provides a cross-cutting discussion. Finally, Section 7 concludes the survey and outlines future directions.

2 Visual Interaction Techniques in Data-Intensive Interfaces

As data volumes and dimensionality grow, interactive visualizations must scale while preserving clarity, responsiveness, and semantic coherence. This section examines techniques that enable users to effectively navigate, interpret, and manipulate large-scale data through visual and interaction-driven mechanisms.

2.1 Scalable Abstractions and Density-Preserving Visual Encoding

Scalable visualization in Big Data systems requires techniques that preserve structural and statistical properties while minimizing visual overload. Direct rendering of raw data quickly leads to overplotting, aliasing, and perceptual saturation. Methods such as kernel density estimation, hexagonal binning, and adaptive sampling summarize dense regions while maintaining statistical variance, enabling users to detect macro-patterns—clusters, gradients, anomalies—without obscuring fine-grained structures [60].

High-dimensional data introduces additional challenges, as visual encoding must project complex manifolds into limited 2D or 3D space. Techniques like principal component analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE), and uniform manifold approximation and projection (UMAP) facilitate this but often introduce distortions and reduce interpretability under interaction. To mitigate these effects, hybrid methods now embed semantic overlays—glyphs, colour encodings, or domain cues—within reduced-dimensional plots, preserving both geometry and context for richer interpretation [18, 33].

Maintaining density-aware representation across zoom levels further complicates visualization. Multiscale frameworks, such as Nanocubes and imMens, enable seamless transitions between overviews and detail by using pre-aggregated spatial-temporal tiles. These are not merely animated transitions but computed recompositions that reaggregate data on-the-fly, requiring precise interpolation strategies to prevent structural artefacts or misrepresentation [29, 30].

2.2 Real-Time Interaction Under Latency Constraints

Interactive analysis of large-scale datasets depends on maintaining a tight perceptual loop between user actions and system feedback. While users expect latency below 200 milliseconds, operations on distributed or disk-resident data frequently exceed this threshold. To bridge this gap, latency-aware interfaces employ progressive, approximate, and predictive strategies that decouple interactivity from full computation [4, 54].

Progressive computation delivers immediate partial results—refining them as more data is processed—making it effective for exploratory tasks like clustering or top-k queries. Systems such as Profiler and SampleAction dynamically tune computation precision based on interaction speed and viewport changes, prioritizing responsiveness over exactness [51].

Asynchronous designs further buffer interaction from backend delays through task queues, speculative execution, and non-blocking rendering. Interaction-aware schedulers use behavioral logs to prefetch likely following states, optimizing perceived latency. These approaches tightly couple frontend event models with backend execution graphs to maintain interface fluidity under load [78].

To preserve interpretability, latency-tolerant systems also integrate uncertainty visualization techniques—such as blur or confidence shading—that communicate the provisional status of results. These cues help manage user expectations and support iterative reasoning even under incomplete computation, ensuring functional continuity and cognitive alignment [55].

2.3 Multi-Modal and Multi-View Exploration Interfaces

In exploratory analysis of heterogeneous datasets, no single view suffices to represent the underlying semantics. Multi-view systems address this by rendering coordinated perspectives—temporal, spatial, categorical—and synchronizing interactions across them. Shared interaction models and consistent synchronization ensure that filters, selections, and highlights propagate meaningfully between views [41, 52].

Coordination can be tight or loose: tight coupling links views directly (e.g., scatterplot–table selection), while loose coupling allows more interpretive flexibility. Both require interoperable data schemas and consistent interaction grammars, especially in modular or web-based environments [58].

Multi-modal integration further complicates coordination across time-series, logs, and text. Issues like desynchronization, encoding mismatches, and semantic divergence are addressed through time warping, co-indexing, or embedding alignment. Mediation layers help unify semantics for cross-type exploration [27].

Adaptive view composition enhances usability by dynamically reorganizing layout and visibility according to user intent. Techniques such as pruning, rearrangement, and attention-aware prioritization optimize space and focus, often guided by interaction history, task context, or gaze-based inference [50].

Table 1 summarizes principal categories of visual interaction techniques that address the challenges of scalability, latency, and interpretability in Big Data environments. Each strategy integrates computational optimization and adaptive interface design to enable efficient exploration, preserve semantic continuity across interaction scales, and support real-world decision-making under high-volume, high-velocity data streams, as illustrated through practical use cases.

Table 1. Visual Interaction Techniques for Scalable and Responsive Big Data Interfaces with Example Use Cases.

Visualization Strategy	**Core Methods**	**Interaction Purpose**	**Design and Performance Criteria**	**Example Use Cases**
Scalable Encoding	Hex binning, PCA, UMAP, t-SNE with semantic overlays	Reduce clutter and preserve structure in dense or high-dimensional data	Clarity, pattern visibility, structural preservation	Large-scale fraud detection; High-dimensional genomics analysis
Latency-Aware Interaction	Progressive rendering, approximate queries, speculative prefetching	Maintain responsiveness during computation or data retrieval	Response time, perceived fluidity, early-stage accuracy	Real-time traffic monitoring; Interactive recommendation exploration
Multi-View Coordination	Brushing/linking, shared filters, synchronized selections	Enable cross-view exploration of diverse attributes	Task efficiency, accuracy, coordination stability	Urban mobility analytics; Cross-linked sales and demographics exploration
Multi-Modal Integration	Time alignment, co-indexing, semantic mappings across modalities	Unify exploration across text, spatial, and temporal data	Interpretive consistency, integration accuracy	Patient health record timelines; Cybersecurity incident investigations
Multiscale Navigation	Pre-aggregated tiles, hierarchical data cubes	Seamlessly move between overview and detail	Zoom latency, semantic continuity, interaction smoothness	Exploration of satellite imagery; Temporal trends in social media analysis

3 Human-Centered System Architectures for Big Data Analytics

Designing interactive Big Data systems requires architectural models that prioritize responsiveness, flexibility, and user-centric data access. This section explores system-level strategies that support dynamic queries, distributed processing, and synchronized interface behavior.

3.1 Adaptive Query Pipelines for Interaction-Driven Workflows

Traditional Big Data pipelines are optimized for static queries and batch throughput, but exploratory analytics demand adaptive pipelines capable of responding to dynamic, user-driven changes in filters, parameters, and aggregations. These pipelines must evolve in real time to accommodate iterative workflows [62].

To achieve this, systems employ fine-grained execution graphs composed of modular operators that can be reordered, rescheduled, or recomputed as the interface state evolves. Strategic materialization and caching localize recomputation, minimizing latency. For example, modifying a filter triggers only partial query re-evaluation, guided by query-aware lineage tracking that preserves transformation dependencies across interface components [32, 74].

Advanced systems further integrate learning-based optimizations. Reinforcement learning (RL) and multi-armed bandit algorithms anticipate query paths by modeling user interaction patterns, dynamically adjusting operator placement and resource allocation. These mechanisms treat interaction as a continuous feedback signal, yielding pipelines that are not only computationally efficient but semantically aligned with evolving analytic intent [56, 69].

3.2 Decentralized Architectures for Scalable Interface Responsiveness

As data infrastructures span cloud, edge, and on-premises nodes, centralized architectures face latency and bandwidth limitations that impair interactivity. To address this, human-centered systems increasingly adopt decentralized architectures that shift computation closer to data sources and UIs. This shift reflects not just a network optimization, but a rethinking of when and where computation should occur to sustain interactive responsiveness [1, 3].

Edge-side processing handles pre-aggregation, filtering, and transformation before transmission, significantly reducing the data volume, especially in high-frequency domains such as sensor analytics and financial systems. Concurrently, client-side computation via WebAssembly or in-browser engines enables lightweight querying and visualization without relying on the server. Together, these decentralisations optimise both data ingress and interaction loops [59, 68, 71].

Distributed coordination is maintained through partition-aware schedulers and the use of eventual consistency models. While strong consistency is vital in transactional systems, exploratory analytics tolerate staleness in favor of responsiveness. Dynamic query routing further adapts to node locality, system load, and user priority, ensuring efficient task allocation. These design patterns collectively enable fluid interaction across geographically dispersed and bandwidth-limited environments [35, 63].

3.3 Middleware for State Synchronization and Interaction Fidelity

Maintaining a consistent interaction state is critical in multi-view, multi-user, and real-time analytics. Middleware serves as the coordination layer between frontends and distributed backends, ensuring state synchronization and managing communication across components [8].

Modern middleware relies on reactive dataflows, where user actions (e.g., selections, transformations) trigger event propagation to shared state stores. To maintain correctness under concurrency and latency, techniques like operational transformation and conflict-free replicated data types (CRDTs) handle rollback, ordering, and resolution [26].

Beyond synchronization, middleware logs interactions, task history, and analytic provenance, supporting session replay, user modeling, and layout reuse. Scalability is achieved through event partitioning, pub-sub messaging, and schema abstraction, enabling modular and extensible systems that support dynamic dashboards, plugin visualizations, and cross-platform interaction [31, 72].

3.4 Interface-Centric Data Shaping and Schema Flexibility

Rigid schemas and static data models pose a significant barrier to fluid interaction in Big Data systems, especially in exploratory settings where user queries evolve dynamically. Traditional schema-first approaches require predefined structures and aggregations, limiting adaptability. Interface-centric data shaping overcomes this by treating the UI as a primary consumer of data, dynamically adapting schemas based on visualization context and user interaction state [13, 67].

Central to this paradigm is schema-on-read, which enables flexible interpretation of semi-structured or nested data without enforcing fixed relational models. Systems like Apache Drill and Presto support on-the-fly schema inference and polymorphic views—such as pivot tables, graph traversals, or JSON (JavaScript Object Notation) expansion—directly aligned with interactive components [7, 45].

Responsiveness is further enhanced through dynamic view generation, which utilizes user-defined lenses to reshape or annotate data streams during rendering. These lenses operate within just-in-time (JIT) query planners optimized for interface responsiveness rather than full backend coverage. Declarative UI grammars, like Vega or Altair, specify the data structure required by each visualization, ensuring semantic alignment with user intent [15, 42].

To support cross-source integration, interface-centric architectures include schema mapping layers or ontology mediators that reconcile naming conflicts, data types, and granularity levels. These mediators enable the seamless merging of heterogeneous datasets without the need for extensive extract, transform, load (ETL) or manual schema engineering, allowing for real-time exploration across previously incompatible domains [5, 21].

Table 2 summarizes key architectural components that enable responsive, adaptive, and consistent interaction in Big Data environments. Each module integrates backend computation with user-facing interfaces, balancing latency, scalability, and dynamic data alignment to sustain fluid exploratory workflows.

Table 2. Human-Centered Architectural Components for Interactive Big Data Systems with Representative Frameworks.

Architectural Module	**Technical Mechanisms**	**Role in Interaction**	**Evaluation and Integration Factors**	**Representative Frameworks**
Adaptive Query Pipelines	Incremental execution, operator caching, lineage tracking, RL planners	Dynamic recomputation based on interaction changes	Latency, recomputation scope, UI reactivity	Apache Drill, DataPolaris
Decentralized Execution	Edge/cloud workflows, WebAssembly, partition-aware routing	Reduce delay and balance load across sources	Round-trip time, throughput, coordination effort	Apache Druid, EdgeDB
Middleware Synchronization	Reactive dataflows, operational transformation, CRDTs, pub-sub models	Maintain consistent interaction state across views or users	Update latency, session coherence, modularity	Firebase, Fluid Framework
Interface-Centric Shaping	Schema-on-read, JIT views, semantic mediation	Align data access with interface and task needs	Query success rate, adaptation speed, maintenance cost	Presto, Drill Fluid Query

4 Cognitive Adaptation and User Modeling in Interactive Big Data Systems

As users engage with complex data environments, systems must adapt to the cognitive demands and evolving analytical behavior. This section examines techniques for inferring user intent, detecting cognitive load, and personalising interfaces to support effective and sustained interaction.

4.1 Intent Inference and Interaction Profiling

Understanding user intent in exploratory analytics is fundamentally different from traditional retrieval tasks. Users often begin without fixed goals and refine hypotheses iteratively, making intent difficult to capture explicitly. Systems must infer intent from behavioral cues such as navigation paths, mouse movement, view changes, and interaction timing [49].

Modern systems utilize probabilistic models derived from interaction logs. Techniques like inverse RL (IRL), Long short-term memory (LSTM), and Transformers detect evolving patterns and focus shifts, producing session-specific profiles that adapt in real time [46, 76].

Intent modeling supports more than passive analysis—it enables interfaces to behave as proactive collaborators. Inferred goals guide data prioritization, ranking, and content preloading. For instance, repeated focus on anomalies can trigger prefetching or visual emphasis on outlier patterns [57].

Importantly, intent inference must remain interpretable and user-correctable. Systems should explain adaptive behaviors and allow users to confirm or revise inferred goals, ensuring alignment with expectations and preserving trust in exploratory workflows [79].

4.2 Adaptive Interface Composition Based on Cognitive Metrics

While intent inference aligns systems with user goals, cognitive adaptation targets the user's ability to process information in real-time. In interactive Big Data environments, cognitive load fluctuates rapidly in response to data complexity, task novelty, and time pressure. Ignoring these dynamics risks user fatigue, stalled decision-making, and analytic errors. Adaptive interfaces respond by regulating visual complexity and interaction tempo based on real-time cognitive signals [17, 44, 53].

Load estimation draws on interaction metrics, such as dwell time, re-query frequency, and click entropy, as well as advanced systems that incorporate physiological inputs, including eye tracking or electroencephalogram (EEG) data. These signals are processed through probabilistic or deep learning models to infer mental effort non-invasively. When overload is detected, the interface adapts by simplifying views, deferring updates, or emphasising relevant content, guided by policies that prioritise usability and cognitive stability [2, 36, 40].

Personalization enhances this process by modeling user-specific thresholds and behaviors over time, integrating both implicit cues and explicit feedback. However, adaptive systems must remain predictable: abrupt or excessive changes can disrupt flow and reduce trust, particularly in collaborative settings with varying cognitive baselines. Effective adaptation thus requires stability-aware strategies that ensure continuity, fairness, and user control [6, 70].

Table 3 outlines key cognitive adaptation strategies that enable interactive Big Data systems to infer user states and personalize interface behavior dynamically. Each approach integrates user modeling techniques to optimize task efficiency, reduce cognitive load, and maintain adaptability across diverse interaction contexts.

Table 3. Cognitive Adaptation and User Modeling in Interactive Big Data Systems with Application Contexts.

Cognitive Focus	Modeling Techniques	Adaptation Goal	Assessment and Deployment Considerations	Application Contexts
Intent Inference	IRL, LSTM behavior models, session embeddings	Anticipate user goals and analytic direction	Prediction accuracy, alignment with task flow	Interactive search assistants, adaptive recommender systems
Cognitive Load Detection	Dwell time, gaze, EEG, zoom-back frequency	Detect mental strain and adapt complexity	Task performance, sensor feasibility, privacy	Real-time learning platforms, adaptive dashboards
Interface Adaptation	View simplification, attention-aware layouts	Regulate interface density based on cognitive state	Usability gains, error reduction, adaptation stability	Context-aware mobile apps, Virtual Reality training environments
Personalization	Interaction history, continuous preference learning	Tailor the interface to the individual and context	Long-term user satisfaction, overfitting risks	Personalized e-commerce interfaces, smart assistants

5 Explainability, Transparency, and Trust in Human-Centered Analytics

As ML models become central to Big Data analytics, interfaces must make their behavior understandable, trustworthy, and accountable. This section explores methods for visual explanation, trust calibration, and auditability that support user comprehension and regulatory alignment.

5.1 Interactive Visual Explanation of Model Decisions

Explainable interfaces convert opaque model behavior into interpretable and interactive representations. In Big Data systems, this requires both global logic and local decision paths to be embedded in dynamic workflows. Visual methods—such as saliency maps, LIME (Local Interpretable Model-agnostic Explanations) and SHAP (SHapley Addictive exPlanations) plots, and decision path overlays—are core tools for analytical reasoning, not peripheral aids [38, 48].

Unlike static outputs, interactive explainers support what-if analysis, allowing users to explore input variations and model sensitivity in real-time. Altering patient attributes in a clinical profile may have direct and indirect effects on a risk score, which is critical in domains such as healthcare or finance [11].

Modern systems increasingly combine visuals with domain-aware narrative scaffolds, delivering explanations that are tailored to different expertise levels. These hybrid interfaces bridge perceptual intuition and semantic clarity [19].

To remain usable, explanation tools must manage complexity. Multi-layered designs reveal information incrementally, adapt to user intent and expertise, and maintain a balance between depth and cognitive load, essential for transparent, scalable AI systems [65].

5.2 Trust Calibration Through Selective Transparency

Trust in analytic systems is dynamic, evolving through interaction based on system behavior, task complexity, and user expertise. Users must rely on system outputs when appropriate, but remain critical in the face of uncertainty or bias. Selective transparency supports this by exposing model rationale and confidence only when relevant, balancing trust with vigilance [9, 39].

Calibration often begins with the visualisation of uncertainty, using cues such as colour gradients or shaded boundaries to frame outputs probabilistically. This helps users recognise prediction limits and avoid overinterpreting results [77].

Transparency should adapt to context: high-stakes tasks require detailed explanations (e.g., feature importance, ambiguity handling), whereas exploratory scenarios may benefit from more concise summaries. Systems must adjust disclosure depth based on user behavior, task sensitivity, and cognitive load [25].

A key challenge is automation bias, where overconfidence arises from overly assertive outputs. Solutions include confidence disclaimers, model override options, and visibility into model provenance—all of which help users form accurate mental models of reliability [66].

Ultimately, trust calibration must be evaluated through outcomes: improved decision quality, increased user engagement, and appropriate scepticism. Trust matters most when it reinforces system validity and fosters responsible human-AI collaboration [47].

5.3 Auditable and Compliant Interaction Logging

In regulated domains, explainability must extend beyond user understanding to support oversight, traceability, and legal accountability. This requires tamper-proof logs that record user actions, model outputs, and data access in verifiable, auditable formats. These logs serve not only as evidence but as interactive tools for reconstructing workflows, validating compliance, and contextualising decisions [24, 37].

Technically, auditability depends on immutable structures, such as Merkle trees, blockchain, or append-only databases, that ensure cryptographic integrity and non-repudiation. This is particularly critical in sensitive sectors such as healthcare or finance, where disputes may arise from specific analytical decisions [22, 64].

Interfaces present these logs via tailored views: timeline replays for analysts, policy reports for compliance teams, and execution traces for developers. Modern frameworks increasingly embed automated checks that flag regulatory or ethical violations, such as biased model behaviour or unauthorised data use [75].

Notably, auditability must support both algorithmic and human interpretation. Without intelligible interfaces, raw logs fail to ensure transparency. Co-designing audit and explanation layers is essential to make Big Data systems not just functional but trustworthy and accountable [61].

Table 4 summarizes principal mechanisms designed to foster model transparency, user trust, and ethical accountability in Big Data systems. Each mechanism is characterized by its interaction objectives, evaluation criteria, and integration with practical toolkits that support interpretable and compliant user experiences.

Table 4. Explainability, Transparency, and Trust Mechanisms in Big Data Interfaces with Supporting Toolkits.

Trust-Enabling Mechanism	**Interface Approaches**	**User-Oriented Function**	**Evaluation and Ethical Implications**	**Supporting Toolkits**
Visual Explanations	SHAP, LIME, saliency maps, what-if tools	Clarify model behavior and decision logic	Explanation accuracy, user insight, cognitive load	SHAP Python Library, Google What-If Tool
Trust Calibration	Confidence previews, adaptive transparency, uncertainty cues	Align user trust with system reliability	Trust scores, rejection rates, overconfidence risk	TrustyAI Toolkit, TensorFlow Uncertainty Estimation
Auditability and Compliance	Immutable logs, execution tracing, policy alerts	Ensure traceability and legal accountability	Audit completeness, usability, regulatory fit	Hyperledger Fabric, Blockchain Auditing Frameworks
Multi-modal Delivery	Visual + textual narratives, domain-specific overlays	Support diverse users and tasks through explanation flexibility	Comprehension rate, domain alignment, clarity	IBM AI Explainability 360, ELI5

6 Discussion

The analysis presented in this survey highlights the evolving convergence of HCI and Big Data systems while uncovering persistent gaps in the integration of interface responsiveness, cognitive adaptation, and trust-aware mechanisms. Although significant advances have been made in individual areas, such as visual interaction, system architectures, cognitive modeling, and explainability, the coherent orchestration of these dimensions into adaptive, human-centered systems remains largely underexplored.

Existing surveys (Table 5) predominantly frame Big Data challenges in terms of infrastructure, storage, and batch analytics. For instance, [16] emphasizes the scalability of data acquisition and processing, without addressing the real-time needs of user interaction. Similarly, [14] focuses on visualization technologies but omits architectural strategies for sustaining latency resilience or mitigating cognitive overload during exploration. In contrast, the present work repositions the user as an active agent, synthesizing advances that enable real-time, interpretable, and cognitively adaptive interfaces for data-intensive environments.

While [43] introduced the conceptual foundation of Human-Data Interaction (HDI), the discussion remained largely theoretical. This survey operationalizes HDI principles by grounding them into deployable techniques such as progressive intent modeling, latency-tolerant multiscale visualizations, and semantic synchronization across heterogeneous modalities. This technical translation addresses a critical gap absent in prior works.

Moreover, whereas surveys like [10] and [12] concentrate on visual scalability in specific contexts (large displays, Linked Data exploration), they fall short in addressing broader systemic challenges, such as cross-modal consistency and cognitive workload regulation. Our contribution explicitly expands the scope to include multi-modal coordination, adaptive system behavior, and trust calibration, all of which are embedded directly into the interaction loop.

The work of [20] brings explainability and ethics into the discussion of ML-driven Big Data systems; however, it treats explainability as an external analytic function rather than an embedded, user-driven interaction capability. In contrast, this survey focuses on how explainability mechanisms must be natively integrated within interactive systems to sustain informed decision-making and ethical compliance under real-time conditions.

Collectively, the findings advocate a paradigm shift: from Big Data systems that prioritize throughput and passive visualization to systems that actively anticipate user intent, dynamically adapt cognitive complexity, and maintain explainability as a continuous, user-governed process. Future Big Data interfaces must integrate architectural flexibility, mental awareness, and ethical transparency not as auxiliary features, but as intrinsic, foundational design principles critical for sustainable, human-centered intelligence.

Table 5. Survey Articles on Human-Computer Interaction in the Big Data Era.

Ref.	Description
[16]	A survey on Big Data covering definitions, applications in business, society, and science, and challenges such as data capture, analysis, and visualization. Emphasises the rise of data-intensive science as a new research paradigm.
[43]	Introduces HDI as a new interdisciplinary field. Focuses on user agency, legibility, and negotiability in data systems. Emphasizes ethical and societal concerns over data use.
[14]	Surveys Big Data visualisation tools, emphasising the strategic role of visualisation in analytics. Discusses functional and non-functional characteristics of primary tools and their challenges with scale and heterogeneity.
[20]	Explores the integration of ML with Big Data, highlighting applications, challenges (e.g., scalability, privacy), and future directions. Stresses ethical implications and interpretable models.
[12]	It focuses on visual exploration techniques on the web of Linked Data. Analyzes state-of-the-art methods and identifies scalability and user interaction as key challenges in large-scale semantic datasets.
[10]	Surveys interactive visualisation on large, high-resolution displays. Covers benefits, interaction techniques, empirical findings, and applications, highlighting challenges in collaborative data analysis at scale.

7 Conclusion

As Big Data systems increasingly drive critical decisions, HCI plays a key role in ensuring transparency, adaptability, and cognitive alignment. This survey examined four interconnected dimensions—scalable visual interaction, responsive architectures, cognitive modelling, and explainability—that link system performance with human reasoning and trust.

Visual methods address scale through progressive rendering and abstraction, while adaptive architectures ensure responsiveness via decentralisation and flexible schemas. Cognitive models infer intent and manage complexity, and trust mechanisms reveal model behaviour through interactive, auditable explanations. Together, these elements form the basis for truly human-centred Big Data systems.

Future research must integrate real-time cognitive signals (e.g., gaze, speech, physiology) into scalable, privacy-preserving adaptive systems. Collaborative analytics will require user models that adjust to shared and divergent goals. Explainability should evolve into domain-aware, long-term trust frameworks that are tailored to specific domains. As generative models advance, systems must strike a balance between user agency and algorithmic control. Addressing these issues will define the next wave of cognitively and ethically aware HCI systems.

References

[1] Ibrar Afzal, Zulfiqar Ahmad, Abdulmohsen Algarni, et al. 2025. A Latency-Aware and Fault-Tolerant Framework for Resource Scheduling and Data Management in Fog-Enabled Smart City Transportation Systems. *Computers, Materials & Continua* 82, 1 (2025).

[2] Muneeb Imtiaz Ahmad, Ingo Keller, David A Robb, and Katrin S Lohan. 2023. A framework to estimate cognitive load using physiological data. *Personal and Ubiquitous Computing* (2023), 1–15.

[3] Siavash M Alamouti, Fay Arjomandi, and Michel Burger. 2022. Hybrid edge cloud: A pragmatic approach for decentralized cloud computing. *IEEE Communications Magazine* 60, 9 (2022), 16–29.

[4] Ioannis Arapakis, Souneil Park, and Martin Pielot. 2021. Impact of response latency on user behaviour in mobile web search. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*. 279–283.

[5] Muhammad Asfand-E-Yar and Ramis Ali. 2020. Semantic integration of heterogeneous databases of same domain using ontology. *IEEE Access* 8 (2020), 77903–77919.

[6] Mirjam Augstein, Eelco Herder, and Wolfgang Wörndl. 2023. *Personalized human-computer interaction*. Walter de Gruyter GmbH & Co KG.

[7] Nelly Barret. 2024. *User-oriented exploration of semi-structured datasets*. Ph. D. Dissertation. Institut Polytechnique de Paris.

[8] Fábio Barreto, Raphael S de Abreu, Marina IP Josué, Eyre Brasil B Montevecchi, Pedro Alves Valentim, and Débora C Muchaluat-Saade. 2023. Providing multimodal and multi-user interactions for digital tv applications. *Multimedia Tools and Applications* 82, 4 (2023), 4821–4846.

[9] Sarah Bayer, Henner Gimpel, and Moritz Markgraf. 2022. The role of domain expertise in trusting and following explainable AI decision support systems. *Journal of Decision Systems* 32, 1 (2022), 110–138.

[10] Ilyasse Belkacem, Christian Tominski, Nicolas Médoc, Søren Knudsen, Raimund Dachselt, and Mohammad Ghoniem. 2024. Interactive Visualization on Large High-Resolution Displays: A Survey. In *Computer Graphics Forum*, Vol. 43. Wiley Online Library, e15001.

[11] Aditya Bhattacharya, Jeroen Ooge, Gregor Stiglic, and Katrien Verbert. 2023. Directive explanations for monitoring the risk of diabetes onset: introducing directive data-centric explanations and combinations to support what-if explorations. In *Proceedings of the 28th international conference on intelligent user interfaces*. 204–219.

[12] Nikos Bikakis and Timos Sellis. 2016. Exploration and visualization in the web of big linked data: A survey of the state of the art. *arXiv preprint arXiv:1601.08059* (2016).

[13] Sandro Bimonte, Yassine Hifdi, Mohammed Maliari, Patrick Marcel, and Stefano Rizzi. 2020. To each his own: Accommodating data variety by a multimodel star schema. In *Proceedings of the 22nd International Workshop on Design, Optimization, Languages and Analytical Processing of Big Data co-located with EDBT/ICDT 2020 Joint Conference (EDBT/ICDT 2020)*.

[14] Enrico G Caldarola and Antonio M Rinaldi. 2017. Big data visualization tools: a survey. *Research Gate* (2017).

[15] Damianos Chatziantoniou, Verena Kantere, R Lukyanenko, BM Samuel, and A Sturm. 2021. Just-In-Time Modeling with DataMingler.. In *ER Demos/Posters*. 43–48.

[16] CL Philip Chen and Chun-Yang Zhang. 2014. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information sciences* 275 (2014), 314–347.

[17] Francesco Chiossi, Johannes Zagermann, Jakob Karolus, Nils Rodrigues, Priscilla Balestrucci, Daniel Weiskopf, Benedikt Ehinger, Tiare Feuchtner, Harald Reiterer, Lewis L Chuang, et al. 2022. Adapting visualizations and interfaces to the user. *it-Information Technology* 64, 4-5 (2022), 133–143.

[18] Raghavendra M Devadas. 2024. 3D visualization using t-SNE dimensionality reduction technique. *Interactive and Dynamic Dashboard: Design Principles* (2024), 119.

[19] Damiano Distante, Stefano Faralli, Steve Rittinghaus, Paolo Rosso, and Nima Samsami. 2022. Domainsenticnet: An ontology and a methodology enabling domain-aware sentic computing. *Cognitive computation* 14, 1 (2022), 62–77.

[20] Elias Dritsas and Maria Trigka. 2025. Exploring the Intersection of Machine Learning and Big Data: A Survey. *Machine Learning and Knowledge Extraction* 7, 1 (2025), 13.

[21] Muhammad Fahad. 2023. Ontology-based Mediation with Quality Criteria. In *International Conference on Business Intelligence*. Springer, 74–88.

[22] Walid Fdhila, Nicholas Stifter, and Aljosha Judmayer. 2022. Challenges and opportunities of blockchain for auditable processes in the healthcare sector. In *International Conference on Business Process Management*. Springer, 68–83.

[23] Meredith Fischer, Nadia Safaeinili, Marie C Haverfield, Cati G Brown-Johnson, Dani Zionts, and Donna M Zulman. 2021. Approach to human-centered, evidence-driven adaptive design (AHEAD) for health care interventions: a proposed framework. *Journal of General Internal Medicine* 36 (2021), 1041–1048.

[24] Patrick Loic Foalem, Leuson Da Silva, Foutse Khomh, Heng Li, and Ettore Merlo. 2025. Logging requirement for continuous auditing of responsible machine learning-based applications. *Empirical Software Engineering* 30, 3 (2025), 97.

[25] Stephen Fox and Vitor Fortes Rey. 2024. A Cognitive Load Theory (CLT) Analysis of Machine Learning Explainability, Transparency, Interpretability, and Shared Interpretability. *Machine Learning & Knowledge Extraction* 6, 3 (2024).

[26] Valentin Goronjic. 2024. *CRDT-based Serverless Middleware for Stateful Objects in the Edge-Cloud Continuum*. Ph. D. Dissertation. Technische Universität Wien.

[27] Alina Gregori, Federica Amici, Ingmar Brilmayer, Aleksandra Ćwiek, Lennart Fritzsche, Susanne Fuchs, Alexander Henlein, Oliver Herbort, Frank Kügler, Jens Lemanski, et al. 2023. A roadmap for technological innovation in multimodal communication research. In *International conference on human-computer interaction*. Springer, 402–438.

[28] Lijie Guo, Elizabeth M Daly, Oznur Alkan, Massimiliano Mattetti, Owen Cornec, and Bart Knijnenburg. 2022. Building trust in interactive machine learning via user contributed interpretable rules. In *Proceedings of the 27th international conference on intelligent user interfaces*. 537–548.

[29] Sarkis Halladjian, David Kouřil, Haichao Miao, M Eduard Gröller, Ivan Viola, and Tobias Isenberg. 2021. Multiscale unfolding: illustratively visualizing the whole genome at a glance. *IEEE Transactions on Visualization and Computer Graphics* 28, 10 (2021), 3456–3470.

[30] Jeffrey Heer and Dominik Moritz. 2023. Mosaic: An architecture for scalable & interoperable data views. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (2023), 436–446.

[31] Stavroula Iatropoulou, Petros Petrou, Sophia Karagiorgou, and Dimitrios Alexandrou. 2021. Towards platform-agnostic and autonomous orchestration of big data services. In *2021 IEEE Seventh International Conference on Big Data Computing Service and Applications (BigDataService)*. IEEE, 1–8.

[32] Max Irendorfer. 2024. *User approaches to knowledge externalization in visual analytics of unstructured data*. Ph. D. Dissertation. Technische Universität Wien.

[33] Dietrich Kammer, Mandy Keck, Thomas Gründer, Alexander Maasch, Thomas Thom, Martin Kleinsteuber, and Rainer Groh. 2020. Glyphboard: Visual exploration of high-dimensional data combining glyphs with dimensionality reduction. *IEEE transactions on visualization and computer graphics* 26, 4 (2020), 1661–1671.

[34] Kuldeep Singh Kaswan, Anupam Baliyan, Jagjit Singh Dhatterwal, and Om Prakash Kaiwartya. 2023. *Big Data Analytics for Human-Computer Interactions: A New Era of Computation: IoT and Big Data Analytics: Volume 3*. Bentham Science Publishers.

[35] Mustafa Ibrahim Khaleel. 2024. A dynamic weight–assignment load balancing approach for workflow scheduling in edge-cloud computing using ameliorated moth flame and rock hyrax optimization algorithms. *Future Generation Computer Systems* 155 (2024), 465–485.

[36] Sean W Kortschot, Greg A Jamieson, and Amrit Prasad. 2022. Detecting and Responding to Information Overload With an Adaptive User Interface. *Human Factors* 64, 4 (2022), 675–693.

[37] Mahesh Vaijainthymala Krishnamoorthy. 2024. Meta-Sealing: A Revolutionizing Integrity Assurance Protocol for Transparent, Tamper-Proof, and Trustworthy AI System. *arXiv preprint arXiv:2411.00069* (2024).

[38] Kostiantyn Kucher, Elmira Zohrevandi, and Carl AL Westin. 2025. Towards Visual Analytics for Explainable AI in Industrial Applications. *Analytics* 4, 1 (2025), 7.

[39] Bingjie Liu. 2021. In AI we trust? Effects of agency locus and transparency on uncertainty reduction in human–AI interaction. *Journal of computer-mediated communication* 26, 6 (2021), 384–402.

[40] Yunmei Liu, Nicolas S Grimaldi, Niosh Basnet, David Wozniak, Eric Chen, Maryam Zahabi, David Kaber, and Jaime Ruiz. 2024. Classifying Cognitive Workload Using Machine Learning Techniques and Non-Intrusive Wearable Devices. In *2024 IEEE 4th International Conference on Human-Machine Systems (ICHMS)*. IEEE, 1–6.

[41] Sehi L'Yi and Nils Gehlenborg. 2022. Multi-view design patterns and responsive visualization for genomics data. *IEEE transactions on visualization and computer graphics* 29, 1 (2022), 559–569.

[42] Andrew M McNutt. 2022. No grammar to rule them all: A survey of json-style dsls for visualization. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2022), 160–170.

[43] Richard Mortier, Hamed Haddadi, Tristan Henderson, Derek McAuley, and Jon Crowcroft. 2014. Human-data interaction: The human face of the data-driven society. *arXiv preprint arXiv:1412.6159* (2014).

[44] Roderick Murray-Smith, John H Williamson, and Sebastian Stein. 2024. Active Inference and Human–Computer Interaction. *arXiv preprint arXiv:2412.14741* (2024).

[45] Hendrik Nolte. 2024. Utilizing HPC Systems to Serve Compute-Intensive Tasks from a Data Lake Managing Sensitive Data. (2024).

[46] Olurotimi Oguntola and Steven Simske. 2023. Context-aware personalization: A systems engineering framework. *Information* 14, 11 (2023), 608.

[47] Kazuo Okamura and Seiji Yamada. 2020. Empirical evaluations of framework for adaptive trust calibration in human-AI cooperation. *IEEE Access* 8 (2020), 220335–220351.

[48] Chandana Panati, Simon Wagner, and Stefan Brüggenwirth. 2022. Feature relevance evaluation using grad-CAM, LIME and SHAP for deep learning SAR data classification. In *2022 23rd International Radar Symposium (IRS)*. IEEE, 457–462.

[49] Veronika Peralta. 2024. *Leveraging users' behavior, intentions and interests for enhancing Exploratory Data Analysis and Data Narration*. Ph. D. Dissertation. Université de Tours.

[50] Ken Pfeuffer, Yasmeen Abdrabou, Augusto Esteves, Radiah Rivu, Yomna Abdelrahman, Stefanie Meitner, Amr Saadi, and Florian Alt. 2021. ARtention: A design space for gaze-adaptive user interfaces in augmented reality. *Computers & Graphics* 95 (2021), 1–12.

[51] Marianne A Procopio. 2021. *Progressive Visualization for Big Data Exploratory Analysis*. Ph. D. Dissertation. Tufts University.

[52] Michail Schwab, David Saffo, Yixuan Zhang, Shash Sinha, Cristina Nita-Rotaru, James Tompkin, Cody Dunne, and Michelle A Borkin. 2020. VisConnect: Distributed event synchronization for collaborative visualization. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2020), 347–357.

[53] Andrew Seyderhelm and Karen Blackmore. 2024. Dynamic Adaptive Surveillance Training in a Virtual Environment Using Real-Time Cognitive Load and Performance. *International Journal of Serious Games* 11, 3 (2024), 109–133.

[54] Zeyuan Shang, Emanuel Zgraggen, Benedetto Buratti, Philipp Eichmann, Navid Karimeddiny, Charlie Meyer, Wesley Runnels, and Tim Kraska. 2021. Davos: a system for interactive data-driven decision making. *Proceedings of the VLDB Endowment* 14, 12 (2021), 2893–2905.

[55] Faizan Siddiqui, Thomas Höllt, and Anna Vilanova. 2021. A progressive approach for uncertainty visualization in diffusion tensor imaging. In *Computer graphics forum*, Vol. 40. Wiley Online Library, 411–422.

[56] Panagiotis Sioulas and Anastasia Ailamaki. 2021. Scalable multi-query execution using reinforcement learning. In *Proceedings of the 2021 International Conference on Management of Data*. 1651–1663.

[57] Hyemi Song, Sai Gopinath, and Zhicheng Liu. 2024. Intents, techniques, and components: a unified analysis of interaction authoring tasks in data visualization. *arXiv preprint arXiv:2409.01399* (2024).

[58] Maoyuan Sun, Akhil Namburi, David Koop, Jian Zhao, Tianyi Li, and Haeyong Chung. 2021. Towards systematic design considerations for visualizing cross-view data relationships. *IEEE Transactions on Visualization and Computer Graphics* 28, 12 (2021), 4741–4756.

[59] Mircea Ţălu. 2025. A Comparative Study of WebAssembly Runtimes: Performance Metrics, Integration Challenges, Application Domains, and Security Features. *Archives of Advanced Engineering Science* (2025), 1–13.

[60] Wenbo Tao, Xinli Hou, Adam Sah, Leilani Battle, Remco Chang, and Michael Stonebraker. 2020. Kyrix-s: Authoring scalable scatterplot visualizations of big data. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2020), 401–411.

[61] Helena Tendedez, Maria-Angela Ferrario, Roisin McNaney, and Adrian Gradinar. 2022. Exploring human-data interaction in clinical decision-making using scenarios: co-design study. *JMIR human factors* 9, 2 (2022), e32456.

[62] Saravanan Thirumuruganathan, Shohedul Hasan, Nick Koudas, and Gautam Das. 2020. Approximate query processing for data exploration using deep generative models. In *2020 IEEE 36th international conference on data engineering (ICDE)*. IEEE, 1309–1320.

[63] B Thiruvenkatam and MB Mukeshkrishnan. 2021. A Novel Replication Protocol Using Scalable Partition Prediction and Information Estimation Algorithm for Improving DCN Data Availability. In *Mobile Computing and Sustainable Informatics: Proceedings of ICMCSI 2021*. Springer, 609–631.

[64] Hui Tian, Jia Wang, Chin-Chen Chang, and Hanyu Quan. 2023. Public auditing of log integrity for shared cloud storage systems via blockchain. *Mobile Networks and Applications* (2023), 1–13.

[65] Allison Timbs. 2025. The Cognitive Convergence of AI and Human Learning: How Learning Theories Illuminate Artificial Intelligence. *Preprint* (2025).

[66] Amy Turner, Meena Kaushik, Mu-Ti Huang, and Srikar Varanasi. 2022. Calibrating trust in AI-assisted decision making. *URL https://api. semanticscholar. org/CorpusID* 235679554 (2022).

[67] Alexandr Vasenev, Wouter Tabingh Suermondt, Atibha Behl, and Johan Lukkien. 2023. Step-wise MBSE introduction into a company: An interface-centric case study. In *2023 18th Annual System of Systems Engineering Conference (SoSe)*. IEEE, 1–6.

[68] Vinod Veeramachaneni. 2025. Edge Computing: Architecture, Applications, and Future Challenges in a Decentralized Era. *Recent Trends in Computer Graphics and Multimedia Technology* 7, 1 (2025), 8–23.

[69] Xinpeng Wang, Songan Zhang, and Huei Peng. 2022. Comprehensive safety evaluation of highly automated vehicles at the roundabout scenario. *IEEE Transactions on Intelligent Transportation Systems* 23, 11 (2022), 20873–20888.

[70] Wei Wu, Youlin Huang, and Lixian Qian. 2024. Social trust and algorithmic equity: The societal perspectives of users' intention to interact with algorithm recommendation systems. *Decision Support Systems* 178 (2024), 114115.

[71] Fatos Xhafa, Burak Kilic, and Paul Krause. 2020. Evaluation of IoT stream processing at edge computing layer for semantic data enrichment. *Future Generation Computer Systems* 105 (2020), 730–736.

[72] Kai Xu, Alvitta Ottley, Conny Walchshofer, Marc Streit, Remco Chang, and John Wenskovitch. 2020. Survey on the analysis of user interactions and visualization provenance. In *Computer Graphics Forum*, Vol. 39. Wiley Online Library, 757–783.

[73] Yongdan Zeng, Ziming Wang, and Wei Zheng. 2023. Research on the Integration of Computer Human-computer Interaction Technology and Multimedia Medical Big Data Interaction System. *Advances in Computer and Communication* 4, 5 (2023), 20.

[74] Yizhuo Zhai, Yu Hao, Zheng Zhang, Weiteng Chen, Guoren Li, Zhiyun Qian, Chengyu Song, Manu Sridharan, Srikanth V Krishnamurthy, Trent Jaeger, et al. 2022. Progressive scrutiny: Incremental detection of ubi bugs in the linux kernel. In *2022 Network and Distributed System Security Symposium*.

[75] Chanyuan Abigail Zhang, Soohyun Cho, and Miklos Vasarhelyi. 2022. Explainable artificial intelligence (XAI) in auditing. *International Journal of Accounting Information Systems* 46 (2022), 100572.

[76] Yaqi Zhang, Xi Wang, Zhiya Zuo, and Dan Fan. 2024. User Behavior Analysis in Online Health Community Based on Inverse Reinforcement Learning. In *Wuhan International Conference on E-business*. Springer, 250–259.

[77] Jieqiong Zhao, Yixuan Wang, Michelle V Mancenido, Erin K Chiou, and Ross Maciejewski. 2023. Evaluating the impact of uncertainty visualization on model reliance. *IEEE Transactions on Visualization and Computer Graphics* (2023).

[78] Yongchun Zheng, Changlong Li, Yi Xiong, Weihong Liu, Cheng Ji, Zongwei Zhu, and Lichen Yu. 2023. iAware: Interaction Aware Task Scheduling for Reducing Resource Contention in Mobile Systems. *ACM Transactions on Embedded Computing Systems* 22, 5s (2023), 1–24.

[79] Liran Zhou, Zhiquan Feng, Hongyue Wang, and Qingbei Guo. 2024. MIUIC: A human-computer collaborative multimodal intention-understanding algorithm incorporating comfort analysis. *International Journal of Human–Computer Interaction* 40, 19 (2024), 6077–6090.