Next Scientists
European International Journal of Multidisciplinary Research and Management Studies

TYPE Original Research
PAGE NO. 54-64

Intent-Aware, Contextual, and Privacy-Conscious Mobility and Recommendation Systems: Towards Robust, Fair, and Scalable Architectures

OPEN ACCESS

SUBMITED 01 October 2025
ACCEPTED 15 October 2025
PUBLISHED 31 October 2025
VOLUME Vol.05 Issue10 2025

Dr. Arman Kulkarni

Department of Computer Science, University of Zurich

Abstract This article presents a comprehensive, theoretically rich synthesis and original conceptual framework that connects research threads from mobility sensing, indoor localization, context-aware computing, anonymity and identity management, recommender systems, user intent modeling, and fairness in algorithmic decision-making. Grounded in a broad set of empirical and theoretical sources, the work advances an integrative research agenda for developing intent-aware systems that are privacy-conscious, resistant to adversarial manipulations, and capable of providing fair and context-sensitive recommendations across mobile and industrial Internet-of-Things (IIoT) environments. The first part of the article revisits the technical foundations of mobile traffic delay estimation and large-scale sensing (Thiagarajan et al., 2009; UC Berkeley/Nokia/NAVTEQ), and then synthesizes this with precise indoor localization techniques (Martin et al., 2010; Kessel & Werner, 2011) to define a layered sensing architecture. The second part integrates privacy and identity management concepts (Ptzmann & Hansen, 2008), describing how pseudonymity and unlinkability can be balanced with intent-aware identity functions to preserve utility while mitigating privacy risks. The third part addresses recommender system advances and fairness concerns (Deldjoo et al., 2023; Gomez-Uribe & Hunt, 2015), and proposes a multi-granularity intent modeling approach that leverages disentangled representations (Dupont, 2018), heterogeneous graph neural networks (Fan et

al., 2019), and contrastive learning approaches to capture sequential user intent (Di, 2022; Guo et al., 2020). Finally, the article articulates an evaluation strategy, potential attack surfaces including relay and NFC-based threats (Roland et al., 2013), limitations, and a detailed research roadmap. Throughout, major assertions are grounded in the cited literature and the discussion offers nuanced counter-arguments and trade-off analyses. The contribution is a cohesive theoretical scaffold intended to guide future empirical work, platform design, and policy-aware engineering efforts for next-generation intent-aware mobility and recommendation systems.

Keywords: Intent modeling; context-aware systems; privacy; indoor localization; recommender systems; fairness; mobility sensing.

INTRODUCTION

The convergence of mobile sensing, indoor localization, context-aware computing, and recommender systems has created opportunities to deliver services that are simultaneously more personalized, situationally relevant, and operationally efficient. Yet this convergence also raises thorny problems: how to model and act on user intent in ways that are accurate and timely; how to preserve user privacy and anonymity while performing identity-dependent functions; how to ensure algorithmic fairness in sequential and session-based recommendations; and how to secure systems against practical threats such as relay attacks and identity spoofing (Thiagarajan et al., 2009; Martin et al., 2010; Ptzmann & Hansen, 2008; Roland et al., 2013; Deldjoo et al., 2023). Each of these concerns has inspired dedicated research strands, but the literature still lacks an integrative theoretical framework that explicitly ties intent modeling to localization, privacy-preserving identity management, and fairness-aware recommendation. This article addresses that gap by assembling a detailed, multi-layered architecture and research agenda that synthesizes and extends established methods and contemporary developments across these domains.

Understanding this gap requires appreciating the distinct but related histories of the contributing areas. Mobility sensing and traffic delay estimation represent an early, successful application of collective mobile sensing—leveraging user devices to build macro-level situational awareness (Thiagarajan et al., 2009; UC Berkeley/Nokia/NAVTEQ). Indoor localization evolved as a parallel need: moving systems and services into enclosed spaces necessitated high-precision location cues derived from phone sensors and signal-processing techniques (Martin et al., 2010; Kessel & Werner, 2011). Context-aware computing provided the conceptual link between raw sensing and actionable service adaptation—arguing that systems should behave differently when embedded in different contexts (Schilit et al., 1994; Adomavicius & Tuzhilin, 2011). Identity and anonymity scholarship articulated the tensions between service personalization and privacy protection—defining concepts such as unlinkability and pseudonymity that later became central to privacy engineering (Ptzmann & Hansen, 2008). Recommender research has evolved from simple collaborative filtering to sophisticated sequential and session-based models that explicitly represent user intent and temporal dynamics (Gomez-Uribe & Hunt, 2015; Guo et al., 2020; Harte et al., 2023), and contemporary discussions foreground fairness, reproducibility, and the limits of neural methods (Dacrema et al., 2019; Deldjoo et al., 2023).

This article argues that the intersection of these threads necessitates a principled, intent-aware architecture that: (1) senses and models user intent across multiple granularities using disentangled and multi-modal representations; (2) preserves privacy through formal identity constructs and operational pseudonymity; (3) maintains fairness and accountability in recommendation outputs; and (4) anticipates and hardens against operational security threats such as relay and contactless payment attacks. The following sections unpack these claims, articulate the proposed architecture and methods in precise detail, and discuss evaluation strategies and open research challenges with careful attention to the trade-offs involved.

METHODOLOGY

The methodological approach of this article is analytic and synthetic: it systematically integrates empirical findings and theoretical constructs from the referenced literature into an extensible conceptual design, and then elucidates method-level

recommendations for modeling, systemization, privacy engineering, and evaluation. The methodology has four complementary strands: conceptual synthesis, architectural decomposition, modeling recommendations, and evaluation and security analysis.

Conceptual synthesis proceeds by mapping core concepts across domains and identifying points of alignment and tension. For example, mobility sensing (Thiagarajan et al., 2009; UC Berkeley/Nokia/NAVTEQ) and indoor localization (Martin et al., 2010; Kessel & Werner, 2011) are mapped into a unified sensing layer with distinct latency, precision, and energy constraints. Context-aware computing provides interpretive layers that mediate between sensing inputs and intent models (Schilit et al., 1994; Adomavicius & Tuzhilin, 2011). Identity and anonymity terms (Ptzmann & Hansen, 2008) are used to annotate the privacy affordances of identity processes at each layer. Recommender system techniques and fairness literature (Gomez-Uribe & Hunt, 2015; Deldjoo et al., 2023; Fan et al., 2019; Di, 2022) are then applied to the intent modeling and decision layers.

Architectural decomposition breaks the system into modular layers and functions: (1) sensing and data ingestion; (2) context aggregation and situational reasoning; (3) intent modeling and representation; (4) identity and privacy management; (5) recommendation and action selection; and (6) monitoring, evaluation, and security response. Each module is detailed below with an emphasis on algorithmic choices, data flows, and privacy/security properties. The decomposition explicitly separates identity functions from intent modeling to allow for privacy-preserving personalization.

Modeling recommendations articulate specific algorithmic strategies for intent detection and sequential recommendation. The article endorses multi-granularity intent units (building on the concept of consecutive intent units in session-based recommendation), disentangled latent representations that separate continuous and discrete intent aspects (Dupont, 2018), heterogeneous graph neural networks for capturing cross-facet connections (Fan et al., 2019), and relay contrastive learning methods for multi-intent dynamics (Di, 2022). These techniques are proposed in combination and with explicit processing steps that account for mobile energy constraints and localization errors (Thiagarajan et al., 2009; Martin et al., 2010).

Evaluation and security analysis specify metrics, experimental designs, and adversarial scenarios. Metrics include standard recommender measures (precision, recall, NDCG), fairness metrics (group-level exposure, disparate impact), privacy metrics (unlinkability probability, re-identification risk), and robustness measures against relay and spoofing attacks (Roland et al., 2013). Experiment designs propose both synthetic lab simulations—where variables can be neatly controlled—and large-scale field trials leveraging crowd-sensing datasets (Thiagarajan et al., 2009; UC Berkeley/Nokia/NAVTEQ).

Throughout, the methodology emphasizes transparency and reproducibility concerns, echoing reproducibility critiques and calling for open benchmarks and detailed reporting of experimental conditions (Gundersen & Kjensmo, 2018; Dacrema et al., 2019).

Architectural Proposal

The proposed architecture is a layered, intent-aware platform designed for deployment in mobile and IIoT contexts where privacy and fairness are key constraints. The architecture is modular by design to facilitate independent verification and to allow alternative components (for example, different intent models or privacy-preserving identity modules) to be swapped without systemic redesign.

Sensing and Data Ingestion Layer. This layer ingests data from mobile devices, network-derived probes, and stationary sensors. The design takes inspiration from large-scale traffic sensing projects that emphasize energy-efficiency and opportunistic sampling (Thiagarajan et al., 2009; UC Berkeley/Nokia/NAVTEQ). Key design points include adaptive sampling policies that regulate sensor usage based on predicted information gain and battery state, and edge pre-processing that transforms raw signals into compressed, privacy-preserving feature vectors

suitable for downstream intent detection. The layer supports both GPS and alternative indoor localization inputs—such as Wi-Fi fingerprinting and inertial sensor fusion—drawing on techniques for precise indoor localization using smartphones (Martin et al., 2010; Kessel & Werner, 2011). The system expects diverse measurement noise profiles: GPS errors, Wi-Fi multipath effects, and IMU drift. To maintain robustness, the ingestion layer attaches uncertainty estimates to location and motion features, enabling probabilistic downstream reasoning.

Context Aggregation and Situational Reasoning Layer. Context is framed as structured, semantically-rich information that augments sensing signals with human-understandable labels (Schilit et al., 1994; Adomavicius & Tuzhilin, 2011). This layer consolidates temporal, spatial, and social context—e.g., time of day, indoor versus outdoor status, group co-location, and device roles. It resolves low-level uncertainty by combining sensor-based probabilistic inference with symbolic rules (e.g., "if motionless for >15 minutes in an indoor location labeled as office, infer work-related context with probability p"). The layer supports context hierarchies that map raw features to higher-level situational states, enabling intent models to reason at multiple temporal and semantic scales.

Intent Modeling and Representation Layer. The intent layer is the analytic core of the architecture. It implements a multi-granularity, multi-modal representation of intent units that captures both immediate and strategic user goals. Drawing upon session-based recommendation research and multi-intent modeling, the design decomposes user behavior into consecutive intent units—a series of short-term goals that are embedded inside longer-term objectives (Guo et al., 2020; Badal Bhushan, 2025). For representation learning, the layer uses disentangled latent spaces to separate continuous behavioral signals (e.g., preferred speed, dwell time) from discrete choices (e.g., category of interest), following methods for learning joint continuous and discrete factors (Dupont, 2018). To capture complex relationships among entities—users, items, locations, device sensors—the layer uses heterogeneous graph neural network constructs that can propagate intent signals across modalities and facets (Fan et al., 2019). For temporal sequencing, the architecture employs multi-policy relay contrastive learning to model multiple concurrent intents and their transitions within sessions (Di, 2022). By combining these approaches, the architecture can represent probabilistic distributions over next-step actions and multi-step goals simultaneously.

Identity and Privacy Management Layer. A core contribution of the design is the explicit decoupling of identity management from intent inference. Identity functions are handled by a privacy-first module that implements pseudonymity, unlinkability, and selective disclosure principles (Ptzmann & Hansen, 2008). The module supports ephemeral pseudonyms tied to limited contexts and scopes—e.g., a pseudonym for "public transit interactions today"—and mechanisms for unlinkable credential use to enable services without global identification. The module's disclosure policies are rule-driven and based on minimal necessary information principles: identity tokens disclose the smallest set of attributes required for a function and are cryptographically bound to prevent replay and relay misuse. The design also proposes local differential privacy-style perturbation for non-critical analytics, carefully balancing utility and privacy risk (Ptzmann & Hansen, 2008; Thiagarajan et al., 2009). Importantly, the identity module provides audit trails for consented disclosures while ensuring that audit metadata itself is privacy-protected.

Recommendation and Action Selection Layer. This layer combines outputs from the intent model and identity module to generate personalized recommendations or control actions. It includes fairness-aware ranking mechanisms, exposure control, and transparency interfaces. The layer uses hybrid recommendation algorithms—mixing collaborative signals, content features, and context-derived intent scores—implemented with careful regularization to avoid overfitting and to preserve reproducibility (Gomez-Uribe & Hunt, 2015; Dacrema et al., 2019). Fairness constraints are applied as soft or hard rules to balance utility and distributional fairness metrics (Deldjoo et al., 2023). For sequential decisions, the architecture provides policy modules that consider multi-step effects of recommendations, avoiding

greedy heuristics that maximize instantaneous click-through at the expense of long-term satisfaction or fairness (Gomez-Uribe & Hunt, 2015).

Monitoring, Evaluation, and Security Response Layer. Continuous monitoring is required to detect drift in intent distributions, privacy violations, and adversarial activity. This layer includes logging (with privacy protections), model performance dashboards, and automated anomaly detection capable of flagging unusual patterns that suggest attacks (e.g., coordinated relay attempts or unusual credential re-use) or degraded model performance. The architecture specifically targets relay-type threats to contactless systems, because prior work has demonstrated these real-world vulnerabilities (Roland et al., 2013). The layer prescribes mitigations including cryptographic session-binding, proximity-attestation protocols, and signal-fingerprinting heuristics to detect relay anomalies when applicable.

Modeling Techniques

This section presents detailed algorithmic recommendations for intent modeling and sequential recommendation within the proposed architecture. The modeling design is intentionally modular, permitting variation and experimentation, but the core suggestions reflect a synthesis of best practices from the literature.

Multi-Granularity Intent Units. The notion of multi-granularity intent units captures the idea that user behavior unfolds on nested timescales: instantaneous micro-intents (e.g., "select item X now"), session-level intents (e.g., "compare cameras for purchase"), and long-term preferences (e.g., "favor budget options over premium") (Guo et al., 2020; Badal Bhushan, 2025). Representing these requires hierarchical models. Practically, a hierarchical recurrent architecture or a transformer with coarse-to-fine positional encodings can encode micro-to-macro transitions. The model should output intent distributions at multiple horizons, allowing downstream ranking policies to trade off immediate click probability with alignment to long-term objectives.

Disentangled Representations. Disentanglement advocates for latent spaces where different dimensions correspond to semantically-meaningful factors (Dupont, 2018). The advantage is interpretability and modularity: if continuous behavioral attributes (like dwell tendencies) and discrete categorical preferences (like interest in "transport" vs "shopping") are disentangled, the system can better generalize and offer controlled interventions (e.g., privacy obfuscation on sensitive dimensions). Implementations can extend variational frameworks that jointly model continuous and discrete latents with dedicated inference networks, ensuring identifiability via suitable inductive biases and regularizations (Dupont, 2018).

Heterogeneous Graph Neural Networks (GNNs). User interactions naturally span heterogeneous entities: users, items, session tokens, locations, and device modalities. Metapath-guided heterogeneous GNNs have been shown to exploit these relationships for intention recommendation by capturing high-order structures and cross-facet dependencies (Fan et al., 2019). GNN layers can propagate intent signals across edges representing temporal adjacency, co-occurrence, or semantic similarity. For instance, a node representing "user session at Location L" can influence item nodes that historically correlate with that location, integrating spatial context into recommendations (Fan et al., 2019).

Relay Contrastive Learning for Multi-Intent Dynamics. Relay contrastive learning is a promising approach for sequential recommendation under multi-intent assumptions (Di, 2022). The method uses contrastive objectives to separate representations of different intents while also learning policies that can "relay" or transition between intents across time. Practically, this involves constructing positive and negative pairs across time windows and learning projection heads that encourage separation of distinct intent modes. This approach is robust against noisy labels and non-stationarity, two common issues in mobile behavioral data (Adams & McGrew, 2017 cited contextually).

Large Language Models (LLMs) as Intent Reasoners. Recent exploratory work suggests LLMs can be leveraged for sequential recommendation tasks by encoding session histories and generating ranked lists

or explanations (Harte et al., 2023). While LLMs offer flexibility and strong contextual reasoning, this article emphasizes a cautious, hybrid approach: use LLMs for interpretive or explanatory tasks and combine them with specialized, efficient models for real-time scoring. This division addresses latency and reproducibility concerns raised in the recommender literature (Dacrema et al., 2019; Gundersen & Kjensmo, 2018).

Fairness and Regularization. Embedding fairness constraints during training can mitigate disparate outcomes across user groups (Deldjoo et al., 2023). These constraints can be applied at the loss level—penalizing divergence in exposure across protected attributes—or at the ranker level—post-processing rank slates to ensure minimum exposure bounds. The article recommends a utility-aware fairness formulation that explicitly models user satisfaction as a latent variable and treats fairness as a distributional regularization rather than a binary filter. Such a method permits nuanced trade-offs and can be tuned to satisfy regulatory or business constraints.

Privacy and Identity Engineering

Privacy preservation starts with precise terminology and principled mechanism design. The consolidated terminology proposed by Ptzmann and Hansen (2008) is a guide: anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management each refer to different properties that must be consciously engineered. The following paragraphs outline concrete mechanisms for achieving these properties in intent-aware systems.

Ephemeral Pseudonyms and Scoped Credentials. Rather than relying on permanent identifiers, systems should issue scoped pseudonyms, cryptographically bound to context-limited credentials (Ptzmann & Hansen, 2008). For example, a transit operator might accept a credential that proves a valid fare without revealing a global identity. Scope-limited credentials reduce linkage potential across contexts, enabling unlinkability: different sessions yield different pseudonyms that cannot be trivially correlated.

Minimal Disclosure and Attribute-Based Credentials. Attribute-based credential systems permit proving properties (e.g., "has monthly subscription") without revealing identity. Combined with blind signature schemes and zero-knowledge proofs, they allow for verifiable statements while preventing unnecessary disclosure (Ptzmann & Hansen, 2008). Implementation complexity is non-trivial, but the privacy gains are substantial.

Local Processing and Noise Injection. Where possible, sensitive inference (e.g., intent classification on highly private dimensions) should be executed locally on devices, with only aggregated or obfuscated signals transmitted. Techniques from local differential privacy—such as randomized response—can apply to telemetry or coarse behavioral summaries before sending to the server. However, the article cautions that utility losses from noise injection must be carefully quantified and that differential privacy mechanisms should be tuned to the expected analysis tasks (Ptzmann & Hansen, 2008).

Auditability and Consent Management. Privacy engineering must include mechanisms for consent that's granular and revocable. Consent logs should themselves be privacy-protected and should enable users to view when and how their data and pseudonyms were used. This layered transparency is essential to maintaining user trust.

Security Considerations and Adversarial Threats

Real-world deployments face adversarial threats that range from data poisoning, inference attacks, to physical-layer relay attacks on near-field systems. The literature documents practical relay attacks on contactless payment systems, demonstrating the feasibility of adversaries bridging NFC transmissions to perform unauthorized transactions (Roland et al., 2013). The proposed architecture integrates multiple defense layers:

Cryptographic Session Binding. Ensure that credentials and ephemeral pseudonyms are cryptographically bound to lifecycle contexts and ephemeral nonces. Session binding reduces replay attacks and prevents straightforward relay amplification unless the attacker also compromises cryptographic elements.

Proximity and Signal Fingerprinting. For proximity-sensitive functions, such as contactless payments or proximity-based authentication, the system should not rely solely on nominal radio presence. Fingerprinting physical-layer properties (timing, multipath characteristics) and combining with device-originated attestation via secure enclaves increases difficulty for relay attackers. Fingerprints are heuristic and may have false positives; thus, they should be combined with policy thresholds and fallback mechanisms (Roland et al., 2013).

Anomaly Detection and Behavioral Profiling. At the monitoring layer, unusual credential usage patterns—like abrupt changes in geographic origin or repeated failed pseudonym binding—should trigger automated risk assessment processes. These processes should reconcile the need for automated defense with fairness and privacy, avoiding overbroad suspensions that unduly affect legitimate users.

Robustness to Label Noise and Non-Stationarity. Mobile sensor data and behavioral logs are noisy and non-stationary. Models trained without accounting for these phenomena may be brittle or susceptible to manipulation (Adams & McGrew, 2017 cited for the context of ML for encrypted traffic). Using robust training techniques, continual learning strategies, and validation on realistic, temporally-distributed datasets all mitigate these risks.

Evaluation Strategy

Robust empirical evaluation requires diverse datasets, clear baselines, and transparent reporting. The recommendations below organize an evaluation approach that is scientifically rigorous and reproducible.

Datasets and Benchmarks. Use a mix of public crowd-sensed datasets for mobility and traffic delay estimation (Thiagarajan et al., 2009; UC Berkeley/Nokia/NAVTEQ) and controlled indoor localization datasets (Martin et al., 2010; Kessel & Werner, 2011) for location fidelity evaluation. For recommender tasks, leverage session-based public benchmarks and generate synthetic sessions that mimic multi-intent behavior using agent-based simulators. Where feasible, collaborate with stakeholders to run field pilots with proper ethics approval and consent.

Metrics. Multiple, complementary metrics are necessary. Recommendation utility should be evaluated with precision, recall, NDCG, and session-level satisfaction proxies (Gomez-Uribe & Hunt, 2015). Fairness metrics must measure exposure parity and disparate impact across protected cohorts (Deldjoo et al., 2023). Privacy risk can be quantified using empirical re-identification experiments and theoretical privacy loss budgets when differential privacy is applied. Localization accuracy is measured via standard distance metrics (mean error) and probabilistic calibration (Martin et al., 2010).

Adversarial Testing. Conduct targeted adversarial experiments: simulate relay attacks on NFC-like protocols to test detection heuristics (Roland et al., 2013); perform membership inference and reconstruction attacks on model outputs to estimate privacy leakage; and inject label noise to analyze robustness (Adams & McGrew, 2017 context). Adversarial testing should not be an afterthought but integrated into the evaluation cycle.

Reproducibility and Open Reporting. Following concerns in the AI and recommender communities, report full experimental details: preprocessing pipelines, hyperparameters, random seeds, and dataset splits (Gundersen & Kjensmo, 2018; Dacrema et al., 2019). Where direct data sharing is impossible due to privacy, provide synthetic replicas or detailed data descriptors to facilitate reproducibility.

Descriptive Results (Analytic Synthesis)

This article is conceptual and synthesizes results from the literature into coherent expectations and empirical hypotheses rather than presenting new field data. The following descriptive results aggregate and extrapolate the likely outcomes when the recommended architecture and modeling techniques are implemented, based on the cited research.

Energy-Aware Sensing Improves Lifetime without Sacrificing Utility. Prior work shows that energy-aware sensing policies can dramatically reduce device power

consumption while maintaining acceptable inference quality when combined with opportunistic sampling and collaborative aggregation (Thiagarajan et al., 2009). Thus, integrating adaptive sampling and edge pre-processing is expected to prolong device operation and lower participation friction in crowd-sensing deployments while preserving sufficient fidelity for intent modeling.

Disentangled Representations Enhance Interpretability and Controlled Interventions. Disentangled representations can facilitate targeted privacy interventions and controlled personalization by isolating sensitive latent dimensions (Dupont, 2018). In practice, these representations should improve the ability to selectively obfuscate or share attributes without collapsing overall model performance. However, disentanglement is not guaranteed; it requires careful model design and appropriate inductive biases.

Heterogeneous GNNs Effectively Capture Cross-Modal Signals. Research indicates that metapath-guided heterogeneous GNNs can capture complex connectivity patterns relevant to intent recommendation tasks (Fan et al., 2019). In mobility and contextual scenarios, graph-based signals like user co-location, shared session membership, and item co-occurrence can enrich intent prediction beyond sequence-only models.

Relay-Contrastive Learning Enhances Sequential Robustness. Relay contrastive learning accounts for multi-intent transitions and appears promising for sequential recommendation, especially under noisy labels (Di, 2022). Combining contrastive objectives with sequential encoders helps models learn more discriminative intent modes and resist short-term perturbations.

Identity Decoupling Preserves Privacy and Maintains Service Utility when Carefully Designed. Techniques such as scoped pseudonyms and attribute-based disclosures enable a pragmatic balance between privacy and service functionality (Ptzmann & Hansen, 2008). The literature suggests that properly engineered, limited-attribute disclosures can allow essential services (e.g., fare validation) without wholesale identity revelation.

Fairness Integration is Necessary but Operationally Complex. Fairness constraints imposed without nuanced understanding of user satisfaction and business constraints can reduce overall utility or introduce unintended biases (Deldjoo et al., 2023). The recommended approach—utility-aware fairness as distributional regularization—aims to provide more flexible and context-sensitive outcomes.

DISCUSSION

This section offers deep interpretation of the architecture and modeling recommendations, discusses limitations and trade-offs, and outlines future directions. The analysis emphasizes the interdependencies among mobility/ localization accuracy, privacy, intent modeling fidelity, and fairness objectives.

Trade-offs Between Privacy and Personalization. Empirical evidence and practical reasoning show that stronger privacy guarantees (e.g., local differential privacy or maximal unlinkability) often reduce the richness of signals available to intent models, affecting personalization accuracy (Ptzmann & Hansen, 2008). The article advocates for a risk-based approach: categorize functions by sensitivity and adopt graded privacy protections. For high-risk attributes, prefer local processing or cryptographic proofs; for low-risk aggregate analytics, consider differentially-private aggregation.

Robustness Versus Responsiveness. Energy-aware sampling and local processing can introduce latency in updating intent models. While energy conservation extends device participation, it may reduce the responsiveness of intent detection to rapid context shifts (Thiagarajan et al., 2009). System designers must balance update frequency with energy constraints, potentially using change-detection triggers to increase sampling when behavior shifts are detected.

Fairness Complexity. Fairness in sequential recommendation is nuanced. Raw parity constraints can penalize minority-tail content unfairly or suppress niche interests. Thus, fairness metrics should be chosen with stakeholder values in mind and should

account for long-term user satisfaction beyond immediate clicks. Empirical evaluation must explore the time dynamics of fairness interventions to ensure they do not degrade long-term engagement or disproportionately affect small user cohorts (Deldjoo et al., 2023).

Adversarial Threat Landscape. The relay attacks documented against NFC systems reveal a need for rigorous operational security design (Roland et al., 2013). However, defenses such as signal fingerprinting may be sensitive to environmental variation and may produce false positives. Robust defense requires layered approaches: cryptography, anomaly detection, and operational policies that enable secure fallbacks.

Reproducibility and Operational Transparency. The AI and recommender communities have raised concerns about reproducibility and the tendency for neural methods to overclaim gains (Dacrema et al., 2019; Gundersen & Kjensmo, 2018). The architecture and research program advocated in this article place reproducibility and open evaluation at the center: publish detailed model descriptions, data generation procedures, and evaluation scripts. Moreover, for deployments, transparency mechanisms—explanations and user-controllable preferences—are essential for trust.

Practical Deployment Considerations. Real-world systems must contend with device heterogeneity, network intermittency, and regulatory constraints. Implementers should design modular systems where different pipelines (e.g., LLM-based explainers vs. efficient on-device scorers) can be swapped depending on deployment constraints. Regulatory compliance for personal data and automated decision-making must be considered early, with legal counsel and ethics review built into project planning.

LIMITATIONS

This article is intentionally synthetic and conceptual rather than empirical. Its conclusions derive from integrating existing studies and projecting likely outcomes for the proposed architecture. There are limitations to this approach:

1. Empirical Validation Required. The practical effectiveness of multi-granularity intent units, disentangled representations, and relay contrastive learning in the precise combinations proposed needs empirical validation across varied contexts and populations.

2. Dataset Bias and Generalization. Public datasets used for benchmarking often do not represent the full diversity of real-world populations and environments. Deployment in different regions, cultures, or device ecologies may present unexpected challenges.

3. Operational Complexity. The proposed privacy and security measures—ephemeral pseudonyms, attribute-based credentials, and layered defenses—add systems complexity and potential usability challenges. Achieving robust, low-friction user experiences while maintaining strong privacy properties will require careful HCI design and iterative user testing.

4. Resource Constraints. High-capacity models (e.g., heterogeneous GNNs, LLM explainers) may have significant computational costs. Ensuring efficiency, especially for on-device components, is necessary for sustainable deployment.

Future Scope

The research agenda emerging from this synthesis includes immediate and long-term directions:

Empirical Field Trials. Conduct controlled field trials in mobility and retail environments to validate energy-aware sensing policies, intent modeling accuracy, and privacy-preserving identity protocols. These trials should include adversarial testbeds for relay and spoofing attacks.

Benchmark Creation. Develop public benchmarks for multi-granularity intent modeling and for fairness-aware sequential recommendation that include realistic mobility and localization errors.

Interpretable Intent Models. Advance techniques for interpreting disentangled latent spaces and for providing succinct reasons for recommendations that are meaningful to end-users.

Policy and Usability Studies. Integrate legal, ethical, and HCI research to design consent mechanisms and privacy disclosures that are usable and effective.

Cross-Domain Transfer. Study transfer learning approaches for applying intent models trained in one domain (e.g., e-commerce) to mobility or IIoT contexts, analyzing the conditions under which transfer preserves fairness and privacy properties.

CONCLUSION

This article offers a rigorous, theoretically-grounded framework for intent-aware, context-sensitive systems that harmonize mobile sensing, indoor localization, privacy-preserving identity management, and fairness-aware recommendation. By integrating multi-granularity intent units, disentangled representations, heterogeneous graph neural modeling, and privacy-first identity mechanisms, the proposed architecture addresses many of the pressing challenges currently fragmenting research across mobility sensing and recommender systems. The work emphasizes layered defenses against adversarial threats, robust evaluation strategies mindful of reproducibility, and the careful trade-offs necessary between personalization and privacy. While empirical validation remains an urgent next step, the conceptual scaffold provided here aims to guide interdisciplinary research, system design, and policy-oriented discussions that will shape the next generation of responsible, effective, intent-aware systems.

REFERENCES

1. Ptzmann, M. Hansen, Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management: a consolidated proposal for terminology, v0.31 (Feb. 2008). URL: http://dud.inf.tudresden.de/Anon_Terminology.shtml

2. Adomavicius G, Tuzhilin A (2011) Context-aware recommender systems. In: Ricci F, Rokach L, Shapira B, Kantor P (eds) Recommender systems handbook. Springer, Berlin, pp 217–256

3. AT&T Alerts, https://alerts.att.com

4. Badal Bhushan. Intent-Aware Identity Management for Autonomous IIoT: A Decentralized, Trust-Driven Security Architecture. International Journal of Computer Applications. 187, 53 (Nov 2025), 30-41. DOI=10.5120/ijca2025925897

5. Bill Schilit, Norman Adams, and Roy Want. Context-aware computing applications. In Proceedings of IEEE Workshop on Mobile Computing Systems and Applications, pages 85-90, Santa Cruz, California, December 1994. IEEE Computer Society Press.

6. Carlos A. Gomez-Uribe and Neil Hunt. 2015. The Netflix Recommender System: Algorithms, Business Value, and Innovation. Transactions on Management Information Systems 6, 4 (2015), 13:1–13:19.

7. Di, Weiqiang. 2022. A multi-intent based multi-policy relay contrastive learning for sequential recommendation. PeerJ Computer Science 8 (Aug. 2022), e1088. https://doi.org/10.7717/peerj-cs.1088

8. Deldjoo, Yashar; Jannach, Dietmar; Bellogin, Alejandro; Difonzo, Alessandro; Zanzonelli, Dario. 2023. Fairness in Recommender Systems: Research Landscape and Future Directions. User Modeling and User-Adapted Interaction 34, 1 (2023), 59–108.

9. Dupont, Emilien. 2018. Learning disentangled joint continuous and discrete representations. In Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS'18). 708–718.

10. Emilien Dupont. 2018. Learning disentangled joint continuous and discrete representations. In Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS'18). 708–718.

11. Gundersen, Odd Erik and Kjensmo, Sigbjørn. 2018. State of the Art: Reproducibility in Artificial

Intelligence. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th Innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18). 1644–1651.

12. Guo, Xueliang; Shi, Chongyang; Liu, Chuanming. 2020. Intention Modeling from Ordered and Unordered Facets for Sequential Recommendation. In Proceedings of The Web Conference 2020. 1127–1137. https://doi.org/10.1145/3366423.3380190

13. Guo, Jiayan; Yang, Yaming; Song, Xiangchen; Zhang, Yuan; Wang, Yujing; Bai, Jing; Zhang, Yan. 2022. Learning Multi-granularity Consecutive User Intent Unit for Session-based Recommendation. In Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining. 343–352. https://doi.org/10.1145/3488560.3498524

14. Harte, Jesse; Zorgdrager, Wouter; Louridas, Panos; Katsifodimos, Asterios; Jannach, Dietmar; Fragkoulis, Marios. 2023. Leveraging Large Language Models for Sequential Recommendation. In 17th ACM Conference on Recommender Systems (Late Breaking Results).

15. J. Bonneau, C. Herley, P. C. van Oorschot and F. Stajano, "The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication."

16. Kessel, M., Werner, M. “SMARTPOS: Accurate and Precise Indoor Positioning on Mobile Phones”. In: The First International Conference on Mobile Services, Resources, and Users (MOBILITY), 2011, pp. 158–163, 2011.

17. Martin, E., Vinyals, O., Friedland, G., and Bajcsy, R. “Precise indoor localization using smart phones,” in International Conference on Multimedia, ser. MM 2010, pp. 787–790.

18. Roland, Michael; Langer, Josef; Scharinger, Josef; Applying relay attacks to Google Wallet. 5th International Workshop on Near Field Communication (NFC), Feb 2013. DOI 10.1109/NFC.2013.6482441

19. Thiagarajan, A. et al. “VTrack: Accurate, Energy-Aware Traffic Delay Estimation Using Mobile Phones,” Proc. 7th ACM SenSys, Berkeley, CA, Nov. 2009.

20. UC Berkeley/Nokia/NAVTEQ, “Mobile Millennium”; http://traffic.berkeley.edu/

21. Maurizio Ferrari Dacrema, Paolo Cremonesi, and Dietmar Jannach. 2019. Are We Really Making Much Progress? A Worrying Analysis of Recent Neural Recommendation Approaches. In Proceedings of the 13th ACM Conference on Recommender Systems (RecSys ’19).

22. Shaohua Fan, Junxiong Zhu, Xiaotian Han, Chuan Shi, Linmei Hu, Biyu Ma, and Yongliang Li. 2019. Metapath-guided Heterogeneous Graph Neural Network for Intent Recommendation. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2478–2486. https://doi.org/10.1145/3292500.3330673

23. Weiqiang Di. 2022. A multi-intent based multi-policy relay contrastive learning for sequential recommendation. PeerJ Computer Science 8 (Aug. 2022), e1088. https://doi.org/10.7717/peerj-cs.1088

24. Yujuan Ding, Yunshan Ma, Wai Keung Wong, and Tat-Seng Chua. 2022. Modeling Instant User Intent and Content-Level Transition for Sequential Fashion Recommendation. IEEE Transactions on Multimedia 24 (2022), 2687–2700. https://doi.org/10.1109/TMM.2021.3088281.