Rethinking Consent for the Age of AI Agents

New Mechanisms for Trust and Transparency

Morgan Rush
Boston, Massachusetts, USA
mleinberry@gmail.com

ABSTRACT

As AI agents increasingly act on behalf of users, existing consent frameworks fail to address two critical scenarios: AI agents accessing personal data to perform tasks, and AI systems sharing user data across business boundaries. This paper discusses mechanisms centered on value-exchange transparency and granular capability controls, exploring consequences of both inaction and potential overcorrection in shaping AI-mediated privacy futures.

1 Consent Crisis

AI agents promise to reshape digital interactions: shopping assistants comparing prices, calendar agents scheduling meetings, health assistants sharing medical data. Yet our consent infrastructure remains rooted in human-to-service interactions where users make discrete, comprehensible choices about data sharing with known entities.

Current mechanisms, such as cookie banners under GDPR, and consent platforms under DMA, assume users understand what data is collected, how it will be used, and with whom it will be shared. AI agents disrupt these assumptions. Users cannot anticipate how AI will use their data. Decisions happen continuously as agents operate autonomously. Data flows across multiple services, creating complex webs that defy traditional disclosure models.

Two challenges emerge: First, AI agents require rich personal data to function effectively, for example booking travel matching preferences, shopping according to dietary needs, managing finances based on spending patterns. Unlike traditional services where users understand functional necessity, AI agents operate with broader, ambiguous data needs. Current mechanisms force all-or-nothing choices that cannot distinguish between "use my purchase history to find deals" and "share my purchase history with marketers."

Second, as AI agents act for users, they share data across organizational boundaries. A shopping agent comparing prices shares browsing patterns with retailers. A scheduling agent reveals calendar availability. A health agent transmits medical information to insurance providers. Users lack mental models for these transactions. When an AI "shops around," is each query a consent event?

Existing frameworks fail because: (1) AI exponentially increases privacy decision complexity creating cognitive overload; (2) point-in-time consent cannot govern continuous AI behavior; (3) users cannot predict how AI will use data given AI opacity; (4) users are told what they surrender but not what they gain.

1.1 From Activities to Outcomes: The New Consent Paradigm

AI-era consent proposes a fundamental shift: from **activity-based permissions** to **outcome-based consent with contextual boundaries**.

The Paradigm Shift: Traditional consent asks permission for activities, such as "Can I access your location?" or "Can I read your purchase history?". Users struggle to evaluate these requests because they cannot predict downstream use. AI agents exacerbate this problem: granting an AI "access to calendar data" provides no meaningful understanding of what the AI will do. The model should be inverted: users consent to outcomes they desire, like "Help me save money on groceries" or "Find the fastest route to work" and these outcomes are transformed into machine-readable contexts that bound AI agent behavior.

Outcome-to-Context Transformation: An outcome-based consent ("save money on groceries") translates into a context specification that AI agents can interpret: permission to access purchase history, compare prices across retailers, suggest alternatives, and share necessary data with e-commerce platforms for price matching but not share data with marketing partners or use data for other purposes. These contexts act as both permissions and constraints, providing AI agents clear operational boundaries while giving users comprehensible control over what AI does on their behalf.

This approach addresses three failures of activity-based consent: (1) **Comprehensibility**: users understand desired outcomes more readily than technical data operations; (2) **Dynamism**: contexts can govern continuous AI behavior rather than requiring per-activity approval; (3) **Value alignment**: consent is organized around user goals rather than system requirements.

Building on this paradigm, there are three mechanisms that can be considered:

Outcome-Based Consent Contexts: Users establish consent by specifying desired outcomes ("save money," "improve health," "save time") and associated constraints (data sharing boundaries, risk tolerances). These declarations are formalized into machine-

The views expressed in this article are solely those of the author and do not necessarily reflect the views of their employer

readable context specifications that AI agents query before actions. A user might consent to outcome "find travel deals" with context constraints: "access calendar and travel history, compare prices with travel sites, do not share data with advertisers." The context serves as both permission and boundary.

Outcome-First Consent Interfaces: When AI systems request new permissions, they present the outcome users will achieve, i.e. "Enable this to save 15% on repeat purchases", before explaining required data access. Interfaces show outcome-for-data exchange using concrete examples ("access your purchase history to identify products you buy regularly and find better prices") rather than abstract legal language. This reframes consent as goal-directed rather than risk-focused.

Context-Aware AI-to-AI Data Sharing Protocols: When AI agents need to share user data across services, a standardized protocol validates sharing against user consent contexts. Before a shopping agent shares browsing patterns with a retailer, the protocol: (a) verifies the sharing serves a consented outcome, (b) checks context constraints (e.g., "no marketing use"), (c) provides real-time notification for high-risk sharing, (d) logs transactions for user review, (e) enforces technical controls (data minimization, purpose limitation) at protocol level. This transforms opaque AI-to-AI data flows into governed transactions aligned with user-specified outcomes.

1.2 Consequences and Alternative Futures

Without Action

Erosion of trust: As users discover AI agents shared data unexpectedly, trust declines. It begins to mirror consent fatigue around cookie banners.

Regulatory backlash: Governments impose blunt regulations, such as blanket prohibitions, mandatory opt-in creating friction, or strict liability discouraging innovation.

Widening digital divide: Sophisticated users opt out, protecting privacy while foregoing benefits; less sophisticated users grant broad permissions without understanding implications. Privacy becomes a luxury good.

Normalization of surveillance: Maximal data collection becomes default as users, overwhelmed by complexity, accept that a lack of control on their data is the price of AI assistance.

With Proposed Approach

Restored agency: Granular, comprehensible controls let users regain meaningful agency, building trust.

Market differentiation: Standardized yet flexible frameworks allow companies to compete on privacy grounds.

Regulatory clarity: Clear mechanisms provide governance templates, reducing legal uncertainty.

Sustainable adoption: When users understand the value exchange, adoption becomes sustainable. Users in control explore capabilities and provide feedback.

Risks of Proposed Approach

Implementation complexity favoring incumbents and harming competition.

Inhibited functionality: If every access requires approval, friction may render agents too cumbersome.

Provocations for Alternative Futures

Consent-Free AI: Could AI agents operate with strict data minimization and purpose limitation enforced technically, eliminating individual consent decisions? Would technical controls better protect users than consent governance?

Collective Consent Governance: Could users delegate consent to trusted collectives that negotiate terms on behalf of members?

AI Consent Assistants: What if users employed personal AI agents to manage consent, fighting AI with AI? This raises questions about trust and potential arms races.

Radical Transparency Without Control: Would perfect visibility into AI data use (i.e. real-time feeds showing every query, access, inference) be more empowering than imperfect control with limited visibility?

1.3 Conclusion

AI agents represent a fundamental shift requiring consent infrastructure evolution. Current mechanisms cannot address AI-mediated data access and AI-to-AI sharing. Consequences of inaction, like eroded trust, regulatory backlash, normalized surveillance, threaten both privacy and AI's beneficial potential.

This proposal is that there is a shift from activity-based permissions to outcome-based consent with contextual boundaries. This paradigm, where users consent to outcomes they desire and these translate into machine-readable contexts governing AI behavior, aims to restore user agency while enabling genuine AI value. There are risks (implementation complexity, inhibited functions) but this approach provides a path forward that aligns consent with user goals rather than system requirements.

Ultimately, AI and consent's future will be shaped by today's design choices. The goal is not consent for its own sake, but empowering users to shape the AI-mediated world they inhabit and providing transparency, tools, and safeguards allowing decisions aligned with user values and visions of a desirable future.

The views expressed in this article are solely those of the author and do not necessarily reflect the views of their employer

Author Biography

Morgan Rush is a technical product manager, focused on consent experiences. Having worked in tech for five+ years, she moved into data privacy over one year ago. She is deeply passionate about improving the customer experience. Outside of work, she is a mother to two young daughters and lives with her husband, daughters, and dog in Weston, Massachusetts.

REFERENCES

[1] Acquisti, A., et al. (2015). Privacy and human behavior in the age of information. *Science*, 347(6221), 509-514.

[2] Barocas, S., & Nissenbaum, H. (2014). Big data's end run around anonymity and consent. *Privacy, Big Data, and the Public Good*. Cambridge.

[3] EDPB. (2020). Guidelines 05/2020 on consent under Regulation 2016/679.

[4] Nissenbaum, H. (2009). *Privacy in Context*. Stanford University Press.

[5] Solove, D. J. (2013). Privacy self-management and the consent dilemma. *Harvard Law Review*, 126, 1880-1903.

The views expressed in this article are solely those of the author and do not necessarily reflect the views of their employer