Ethnography, Design Thinking, and Jobs-to-be-Done in Public-Sector Service Redesign: A Survey of Two Decades of Nordic Practice and Its Translation to LLM-Mediated Virtual Experiments

Akira Sato [1] **and Claude Opus 4.7 (1M context)** [2]

[1] Independent Researcher [2] Anthropic — operating as co-author under Honest Design Principle (a) of explicit epistemic-dependency disclosure [rc-20260427]

Preprint target: aiXiv (https://aixiv.science/) **Date**: 2026-05-20 **Type**: Survey paper (systematic literature review + methodological synthesis) **Version**: v0.2 (survey-paper restructure of v0.1 position paper) **Word count**: ~11,000 words

Abstract

This paper presents a systematic survey of two decades of public-sector service redesign methodology, with particular emphasis on the Nordic tradition (Denmark's MindLab and borger.dk, Estonia's X-Road, Norway's Disruptive Taskforce, Finland's Public Sector Innovation Fund), the UK Government Digital Service (GDS), and Japan's emerging Connected One-Stop initiative. We organize the surveyed material along three intersecting traditions — **ethnographic public-sector research**, **design thinking applied to government**, and **Jobs-to-be-Done (JTBD) for life-event services** — and propose a *translation table* that maps each surveyed methodological primitive to its counterpart in LLM-mediated virtual experiment design. The survey covers 60+ peer-reviewed and institutional sources across 2002–2026, identifies seven persistent methodological gaps, and articulates how the Sato 2026 research series (ARC, 2026v, 2026w, 2027a, 2027b) instantiated the translation. We argue that, in the post-Mythos AI era where expert-level cognition is no longer scarce, what becomes scarce — and therefore valuable — is *disciplined methodological practice* of the kind that Nordic service-design traditions have refined for citizens and that LLM-mediated simulation now inherits for virtual experiments.

Keywords: survey, public-sector innovation, ethnography, design thinking, Jobs-to-be-Done, service blueprint, journey mapping, MindLab, borger.dk, X-Road, GDS, LLM-mediated simulation, virtual experiments, methodological closure, recursive limitation.

1. Introduction

1.1 Motivation

The 2024–2026 surge in LLM-mediated simulation — Generative Agents [Park et al. 2023], Sotopia [Zhou et al. 2023], multi-actor neurodivergent triadic simulation [Sato & Claude 2026a] — has outpaced its

methodological grammar. Researchers routinely instantiate LLM "personas" and report aggregate outcomes, but without a disciplined design vocabulary the simulations risk what Bisbee et al. [2024] flagged for synthetic survey replacement: surface-plausible artifacts whose internal logic is unmoored from any lived institutional reality.

A separate body of work — Nordic public-sector service redesign — has, since the early 2000s, developed precisely the vocabulary that LLM simulation lacks: *life-event framing*, *multi-persona modeling*, *service blueprints*, *journey-stage KPIs*, and *continuous ethnographic monitoring*. This vocabulary was developed for redesigning the experience of citizens navigating fragmented government services. It has, to our knowledge, never been systematically surveyed nor systematically translated to LLM-mediated virtual experiment design.

1.2 Survey Objectives

This survey pursues four objectives:

1. **Reconstruct the lineage** of Nordic public-sector ethnographic service-design methodology and place it in dialogue with adjacent traditions (UK GDS, Estonia X-Road, Japan Connected One-Stop, Singapore Co-Delivery).
2. **Distill operational primitives** that recur across the surveyed practice (life-event framing, multi-persona, As-Is/To-Be journey mapping, service blueprinting, KPI frameworks, co-creation workshops).
3. **Identify persistent gaps** in the surveyed literature: cross-organizational service revision, citizen co-creation, neurodivergent accessibility, implementation capacity, and the risk of "automation exclusion" in proactive-government modes.
4. **Articulate a translation** from public-sector service-design primitives to LLM-mediated virtual experiment design — including a concrete translation table (§7.1) and worked applications to the Sato 2026 research series (§8).

1.3 Roadmap and Scope

§2 specifies the survey methodology, scope, and source-selection criteria. §3 surveys the Nordic ethnographic lineage. §4 surveys design thinking applied to government. §5 surveys JTBD in public-sector life-event redesign. §6 surveys KPI frameworks. §7 presents cross-cutting taxonomy, gap analysis, and the translation to LLM virtual experiments. §8 applies the translation to five Sato 2026 research projects. §9 addresses limitations and recursive limitations. §10 articulates open research questions and future work.

The survey *does not* cover: corporate (private-sector) ethnographic design (Suri 2003; IDEO methods, except where directly applied to public-sector); behavioral nudging frameworks (Sunstein 2014; though Behavioral Insights Teams are touched in §4.3); or the broader implementation-science literature on welfare-state digital transformation.

---

2. Survey Methodology

2.1 Source Selection

We assembled the survey corpus by combining:

(a) **Citation-grounded query results** from Perplexity Sonar, executed across the Sato 2026 research series in the preparation of this paper and its predecessors. Approximately 30 queries directly relevant to public-sector ethnography, JTBD, and KPI frameworks contributed to the source list, with all citations verified against primary documents where retrievable.

(b) **Institutional documentation** from public-sector innovation labs (MindLab Denmark, La 27e Région France, Helsinki Lab, Singapore Open Government Products, US 18F, UK GDS), retrieved from official publications between 2002 and 2026.

(c) **Peer-reviewed literature** identified via Google Scholar and the digital archives of journals including *International Journal of Design*, *Sustainable Cities and Society*, *Service Industries Journal*, *Public Administration Review*, and *Journal of Service Research.*

(d) **Grey literature**: OECD reports, UN Development Programme service-design toolkits, KPMG and Boston Consulting Group case studies, OECD Observatory of Public Sector Innovation (OPSI) toolkits.

2.2 Inclusion Criteria

A source was included if it: 1. Documented an actual public-sector service redesign (not a purely theoretical paper); 2. Specified the methodology used (interview protocol, journey-mapping technique, KPI framework, etc.); 3. Was authored either by primary participants in the redesign or by independent researchers analyzing it; and 4. Was retrievable in English, Japanese, or one of the major Nordic languages with English-language summary.

2.3 Exclusion Criteria

Excluded: marketing materials lacking methodological detail; purely architectural / technical specifications without service-design context; sources unable to be cross-verified against at least one independent reference.

2.4 Limitations of the Survey Methodology

This survey is *not* a PRISMA-style systematic review and does not claim exhaustiveness. The Perplexity-grounded retrieval underrepresents non-English non-Japanese non-Scandinavian sources. The survey is also LLM-mediated and is subject to the methodological-closure critique we will return to in §9.

---

3. Nordic Ethnographic Lineage

3.1 MindLab Denmark (2002–2018)

In 2002 the Danish Ministries of Employment, Taxation, and Economic and Business Affairs jointly established **MindLab**, the world's first cross-ministerial public-sector innovation lab [Bason 2010; IAIS 2019; OECD-OPSI 2017]. MindLab's three institutional moves became foundational:

1. **Cross-ministerial mandate**: reporting to three ministries simultaneously, structurally bypassing the siloing that creates ping-pong service experiences.
2. **Ethnographic primacy**: interview, in-situ observation, and journey-map construction *preceded* policy or technical design.
3. **Citizen-vocabulary discipline**: services named in citizens' words — "moving home," "becoming a parent," "starting a business" — never in administrative terminology.

MindLab was closed in 2018 [govinsider 2018] and partially succeeded by a "Disruption Taskforce" focused on implementation. The closure itself is methodologically informative: as Bason and others have argued, the lab's separation from operational ministries eventually became a liability, suggesting that the next-generation public-sector innovation must be *embedded* rather than *adjacent* to delivery [Bason 2018].

3.2 borger.dk and Min Side

Denmark's national citizen portal **borger.dk**, operated by the Agency for Digitalization (DIGST), is the most thoroughly documented application of MindLab-style methodology [Igari 2024; NHK STRL 2019; DIGST documentation]. A 2005 foundational user-research study established the *life-event navigation* principle: citizens want to navigate from their lived situation ("we are expecting a child," "I just lost my job") rather than from administrative categories ("Ministry of Children and Education," "Ministry of Employment").

The signature persona of borger.dk — Klaus, the 44-year-old carpenter in a village near Bjerringbro — is paired with national statistics (e.g., "72% of Danes did online shopping in 2011") to ground each persona in measurable representativeness rather than narrative appeal. Multiple personas (Klaus the offline-leaning rural worker, an elderly user, a non-Danish-speaking resident, a person with cognitive accessibility needs) partition the design space; design decisions are evaluated for whether they serve *each* persona, not merely the convenient majority.

Min Side, the authenticated personal page introduced in 2008 with the NemID national identity system, extended the design discipline to runtime personalization: persona attributes become system variables, enabling proactive service delivery (e.g., automated MOT-renewal reminders for car owners).

3.3 Finland: Public Sector Innovation Fund and "TM"

Finland's Public Sector Innovation Fund and the SITRA-affiliated Helsinki Lab extended the Nordic tradition with two distinctive contributions [SITRA 2019]:

1. **Co-creation workshops with ministerial leadership**: senior officials are required to participate in citizen-design workshops, breaking the lab-versus-delivery separation that eventually closed MindLab.

2. **Behavioral-design integration**: the Helsinki Lab integrated behavioral economics with ethnographic service design, creating a hybrid that explicitly tracked citizen "default settings" and unconscious heuristics.

3.4 Norway: Difi, "DigiHelse," and the OECD Critique

Norway pursued a parallel agenda through Difi (the Agency for Public Management and eGovernment) and the DigiHelse program for digital health services. The 2017 OECD Digital Government Review of Norway [OECD 2017] documents a persistent gap: while Norway's ministries each implemented sophisticated digital services, the lack of inter-ministerial coordination produced "Nordic-style vertical digitalization" — each ministry's portal is excellent on its own; together they re-instantiate the ping-pong problem the methodology was supposed to solve. The OECD critique frames Nordic ethnographic design as necessary but insufficient: cross-ministerial *governance* of the data and process architecture matters as much as cross-ministerial ethnography.

3.5 Estonia: X-Road and the Once-Only Principle

Estonia's contribution is *architectural* rather than ethnographic. X-Road, the inter-agency data exchange backbone (launched 2001, current version 7), made the Once-Only Principle technically realizable: a citizen provides a piece of information once, and government systems exchange it among themselves [Ash Harvard 2024]. Robinson [2020] documents that independent analyses estimate X-Road saves Estonians approximately 844 person-years annually in form-filling.

The methodological lesson is that *ethnography surfaces the experiential signature of bureaucratic ping-pong* (frustration, exclusion, abandonment); *architecture removes its technical substrate*. The two are complementary, not substitutable.

3.6 Cross-Nordic Synthesis

Across the Nordic lineage we identify four shared commitments:

1. **Cross-organizational mandate** (MindLab, Finland's leadership co-creation, X-Road's inter-ministerial protocol) — the recognition that fragmentation at the supply side cannot be hidden from the demand side.
2. **Citizen vocabulary discipline** — services named in life-event terms, not administrative ones.
3. **Multi-persona partition** — the design space partitioned by representative personas, with each persona's serviceability evaluated separately.
4. **Continuous post-deployment ethnography** — borger.dk's three-stage rhythm of co-creation → pilot → continuous feedback represents the maturation of the discipline from one-shot redesign to ongoing accommodation of changing citizen contexts.

---

4. Design Thinking Applied to Government

4.1 The Stanford d.school / IDEO Lineage in Public Sector

While "design thinking" as a popularized methodology originates in IDEO and Stanford d.school [Brown 2009; Kelley & Kelley 2013], its public-sector adoption has been distinctive [Bason 2018; Mortimer 2020]. The five-stage canonical sequence (Empathize → Define → Ideate → Prototype → Test) translates to government with three modifications:

1. **Empathize** in government means *ethnographic immersion*, not user-interview convenience. Public-sector users include people the designer cannot easily contact (housebound elders, undocumented residents, people experiencing homelessness, people in psychiatric care).
2. **Define** must include *jurisdictional scope*: which ministries, which laws, which budgets. Government design without jurisdictional clarity produces designs that no one is empowered to implement.
3. **Prototype** must include *policy prototypes* (proposed regulatory changes, draft statutory language) alongside user-facing prototypes. Public-sector innovation lab Bloomberg Philanthropies' i-teams framework formalizes this dual prototyping.

4.2 The UK Government Digital Service (GDS)

The UK GDS, founded in 2011 [Bracken 2016; Government Digital Service 2014], crystallized design thinking into a *Service Standard* of 24 design principles, mandatory for every major UK digital service. GDS's signature contributions include:

1. **The Four Mandatory KPIs** (Completion Rate, User Satisfaction, Cost per Transaction, Digital Take-up) — covered in §6.
2. **Continuous user research at every life-cycle stage** — alpha, beta, live, and post-live all have mandatory user research checkpoints.
3. **The "Government as a Platform" architecture** [Bracken 2016] — shared platform services (Notify, Pay, Verify) that each individual service uses, reducing duplication and improving consistency.
4. **The blogged-iteration culture** [designnotes.blog.gov.uk] — public documentation of design decisions and their rationale, creating a public-domain methodology repository.

GDS's international diffusion has been substantial: the 2025 Helsinki-Amsterdam international design discourse meeting [GDS 2025] brought together 30+ governments to share GDS-derived methodology.

4.3 Behavioral Insights Teams ("Nudge Units")

Parallel to GDS's design-thinking emphasis, the UK Behavioural Insights Team (now Nesta-affiliated) and its international analogues (US Office of Evaluation Sciences, Australia BETA, EU CORE) developed a complementary methodology focused on small-scale field experiments and behavioral defaults [Halpern 2015]. While "nudge units" are not ethnographic design teams, their RCT-based methodology has been productively combined with GDS-style design thinking in projects like UK welfare-claims auto-completion.

4.4 Canadian and Singapore Adoption

Canada's "Service Lab" and the Canadian Digital Service [TBS 2020] adapted GDS methodology with explicit attention to Indigenous-community service design, including the "Reconciliation by Design" principle. Singapore's Open Government Products (OGP) [GovTech Singapore 2021] adapted the methodology with a Co-Delivery governance pattern: cross-ministerial improvement proposals are automatically routed to all affected agencies, with a designated "journey custodian" responsible for end-to-end outcome.

4.5 Japan's Service Design Twelve Principles

Japan's Digital Agency published its Service Design Twelve Principles (2024-2025) [Digital Agency 2024] in explicit dialogue with GDS and DIGST precedent. The principles include: starting from user needs, designing for inclusion, using data to drive decisions, and continuous improvement. KPMG's 2023 analysis [KPMG 2023] identifies two persistent gaps in Japanese adoption: cross-organizational service revision and citizen co-creation. The neurodiversity-aware extension of this critique [Apreleva 2024; Sato & Claude 2026d] is the subject of §8.5.

---

5. Jobs-to-be-Done in Public-Sector Life-Event Services

5.1 JTBD Foundations

Jobs-to-be-Done theory [Christensen et al. 2016; Ulwick 2017] reframes user needs from *features and demographics* to *life situations and progress sought*. A user does not want a service; they want a *job done*. The theory distinguishes three layers:

1. **Functional jobs**: the practical work to be accomplished (file a tax return, register a child for school).
2. **Emotional jobs**: how the user wants to feel during and after (calm, confident, not anxious or ashamed).
3. **Social jobs**: how the user wants to be perceived by others (a competent adult, a responsible parent, a recognized citizen).

5.2 Public-Sector JTBD: Ulwick (2017)

Ulwick's "Jobs to Be Done for Government" essay [Ulwick 2017] articulates the most direct application of JTBD theory to public services. The reframing makes one move: instead of "use the unemployment benefits service," the JTBD is "regain financial stability after job loss." That move dissolves the ministerial-service framing and reveals the cross-ministerial journey the citizen actually traverses.

5.3 LinkedIn Skill-Development Variant

Brantley's LinkedIn "Government Talent Development: Applying Jobs-to-be-Done Theory to Skill Development" [Brantley 2018] extends JTBD to the workforce-development side: government employees themselves have JTBDs about their professional progression, and treating skill-development services through JTBD logic produces qualitatively different professional-development infrastructures.

5.4 Wicked7 Multi-Stakeholder JTBD

Wicked7's "Multi-Stakeholder JTBD" reformulation [Wicked7 2020] introduces a critical extension: in public-sector contexts, a single life-event affects multiple stakeholders simultaneously (citizen, family, community, employer, agency staff), each with their own JTBD. The design challenge is to articulate a *coherent multi-stakeholder JTBD* — one that does not optimize for the citizen at the cost of the agency staff, or vice versa.

5.5 borger.dk Implementation of Multi-Stakeholder JTBD

borger.dk's life-event structure (e.g., "Having a Baby" combines pre-natal, birth registration, parental leave, child benefit, child healthcare, child registration with municipality, and early-childhood education enrollment) instantiates multi-stakeholder JTBD: the citizen's job is "successful integration of a new family member"; the agency staff's job is "facilitate that integration with minimum friction"; the community's job is "absorb the new family member with appropriate resources." The portal's success depends on serving all three.

5.6 Three-Layer Job-Bug Detection

The Nordic methodology's productive synthesis with JTBD is its three-layer "job-bug" detection: an ethnographic study surfaces *not only functional inefficiencies* but *emotional and social damage* — citizens who avoid services, who feel shame, who are silenced by ping-pong. The journey map records emotional-state inflections (frustration, anxiety, relief) alongside functional steps, creating a multi-dimensional view of where the journey fails.

---

6. KPI Frameworks for Citizen Portals

6.1 UK GDS Four-Mandatory KPI Framework

Codified by the UK Cabinet Office in 2013 [GDS 2013] and mandatory for all major UK digital services since:

KPI	Definition	Failure mode if used alone
Completion Rate	Fraction of users who finish what they started	Drives "fake completion" — narrow definitions of "done"
User Satisfaction	Post-completion satisfaction score	Drives "smile theater" — surface UX polish without substantive improvement
Cost per Transaction	Government-side cost per completed transaction	Drives cost-cutting at expense of citizen experience
Digital Take-up	Fraction of transactions completed via digital channel	Excludes non-digital users (drives digital exclusion)

The framework's discipline is that all four KPIs must be published; the tradeoffs are visible and accountable.

6.2 EU Journey-Stage KPI Framework

The EU Single Digital Gateway evaluation framework [European Commission 2020] articulates KPIs by journey stage:

Access: discovery rate, "don't know where to ask" rate, life-event-page reach, search-keyword distribution.

Navigation: time-to-complete (mean and median), drop-off rate by step, first-contact-resolution rate, form-resubmission rate, error rate by form field.

Achievement: actual service-received rate (not merely application submitted), agency-side SLA performance, rejection rate, days-from-application-to-outcome.

Aftercare: post-completion error correction rate, life-event continuity (e.g., after birth → childcare → school continuity), follow-on service uptake.

6.3 Japanese Adoption Through MIC / DPRO

Japan's Ministry of Internal Affairs and Communications and the Digital Promotion Office (DPRO) have adapted both frameworks for the Japanese context [MIC 2024], with explicit attention to: (i) digital take-up among elderly citizens (>65) as a separate cohort metric; (ii) "no-writing window" digitization rate as a uniquely Japanese KPI; (iii) cross-agency journey completion as a future-priority KPI.

6.4 The "Outcome vs Output" Critique

A recurring methodological critique [O'Reilly et al. 2022; Janssen 2019] is that public-sector KPIs gravitate to outputs (transactions completed, portals visited) and away from outcomes (problems solved, lives improved). Outcome KPIs are harder to measure but methodologically more honest. The Nordic discipline of journey-stage KPIs (Access / Navigation / Achievement / Aftercare) is an attempt to push toward outcome measurement.

6.5 The Equity / Inclusion KPI Frontier

The 2024–2026 frontier of public-sector KPI research focuses on equity and inclusion measurement [Eubanks 2018; Benjamin 2019; Apreleva 2024]. Proposed equity KPIs include: service-uptake rate by demographic subgroup (race, income, disability, language); the "exclusion rate" of automated services (people whom the system fails to recognize or serve); and "trust gain" KPIs (longitudinal change in citizen trust in public institutions).

---

7. Cross-Cutting Taxonomy and Translation to LLM Virtual Experiments

7.1 The Translation Table

From the surveyed material we extract a translation table linking public-sector service-design primitives to their LLM-mediated virtual-experiment counterparts:

Public-sector primitive	LLM virtual-experiment counterpart
Life-event (citizen vocabulary)	Cross-stakeholder JTBD framing of the simulation
Multi-persona (statistically representative)	Multi-LLM-persona partition of the design space
As-Is journey map	Baseline trace of un-augmented LLM simulation
To-Be journey map	Designed intervention with anti-stereotype revision
Service blueprint (front + back stage)	Six-component LLM harness (E/T/C/S/L/V)
Co-creation workshop	Theorist + practitioner + LLM panel + grounded reviewer four-party loop
Continuous ethnographic monitoring	Continuous run-event JSONL audit log + reflexive cycle
UK GDS four KPIs	Adapted: completion / inter-LLM-rater satisfaction / API-cost / take-up
EU journey-stage KPIs	Access / Navigation / Achievement / Aftercare per simulated JTBD
Journey Advisory Board (US pattern)	External web-grounded reviewer (e.g., Perplexity Sonar)
X-Road Once-Only	Append-only event store (E1 pattern), no information re-elicitation
MindLab cross-ministerial mandate	Multi-LLM cross-vendor panel (heterogeneous-vendor adjudication)
Behavioral Insights Team RCT	Cycle 7 protocol participatory review
Equity/inclusion KPI	Anti-stereotype revision cycle improvement score

7.2 Seven Persistent Methodological Gaps

The survey identifies seven gaps that persist across the surveyed practice:

1. **Cross-organizational governance ≠ cross-organizational ethnography**: surfacing the journey is not the same as having the legal-budgetary authority to redesign it [OECD 2017].
2. **The implementation-capability gap**: designers can identify problems; only "internal change agents" embedded in agencies can implement [Bækkelie NTNU thesis 2020].
3. **The neurodiversity gap**: complex procedures × ping-pong compounding falls hardest on users with executive-function differences [Apreleva 2024]; this population has historically been under-represented in the ethnographic record.
4. **The automation-exclusion gap**: proactive-government modes (where the state delivers services without citizen application) systematically exclude users the system cannot recognize (homeless, undocumented, non-typical-life-pattern users) [Eubanks 2018].
5. **The KPI-outcome gap**: published KPIs gravitate to outputs; outcomes are harder to measure and consequently under-tracked [O'Reilly 2022].
6. **The post-deployment monitoring gap**: many projects launch with ethnographic discovery but maintain only minimal post-deployment monitoring [DIGST documentation; SITRA 2019].
7. **The methodological-closure gap** (this paper's contribution): increasingly, the ethnographic, design, and KPI-evaluation work is itself partially LLM-mediated, creating an epistemic-closure problem analogous to the one Bisbee [2024] flagged for synthetic survey replacement.

7.3 What the Translation to LLM Virtual Experiments Does

The translation accomplishes three things:

1. **Vocabulary**: rather than describing an LLM simulation as a "prompt experiment," we describe it as a "virtual journey map of a cross-stakeholder JTBD across an instrumented harness," forcing specification of the persona partition, the back-stage harness, the KPIs, and the continuous-monitoring loop.
2. **Discipline against narrative collapse**: multi-persona modeling, borrowed correctly, prevents the LLM-simulation antipattern of selecting personas for narrative convenience; the discipline is to partition the design space.
3. **Accountability framework**: the GDS four-KPI framework, adapted, forces simulators to publish completion, satisfaction (as inter-LLM-rater agreement), API-cost-per-outcome, and digital-take-up (as LLM-vs-human contribution ratio).

7.4 What the Translation Does *Not* Buy

It does not buy us *external validity*. A virtual experiment, however well-designed, remains an experiment in LLM-simulated experience. The translation gives us better internal coherence and better instrumentation; it does not give us empirical truth. For empirical truth we still need Cycle 7 (real lived-experience review) and the four-party loop's external reviewer (§9.1).

---

8. Worked Applications: Sato 2026 Series

8.1 Sato 2026-ARC: Triadic Neurodivergent Simulation

Life-event framing: "neurodivergent person sustaining grayzone employment over twelve months" with three stakeholders (the person, the support worker, the employer) [Sato & Claude 2026a].

Multi-persona: P1–P8, partitioning along neurotype × employment context × family role × support relationship.

As-Is journey: twelve-month synthetic narrative produced by Claude in v0.2.5.

Service blueprint: arc-rs runner + JSONL event store + Decision events + companion server, realizing the six-component harness.

To-Be journey: anti-stereotype revision cycle in v0.3 — seven stereotype patterns (caretaker support staff, troubled lived-experience holder, unaware employer, burnout, crisis-response trope, predictable adaptive-disorder pattern, vocabulary bleed) independently detected by two LLMs and deliberately inverted, with three independent LLMs converging on a 9/10 improvement score.

KPIs: tier-on-tier multi-LLM review score progression (Tier 1 = 8/10 → Tier 2 = 9/10); Falsification-Criteria status progression (12 of 18 SUPPORTED in v0.2).

Public-sector ancestor: borger.dk's "Having a Baby" life-event with multi-stakeholder JTBD.

8.2 Sato 2026v: Cultivating-AI Natural Experiment

The Sato 2026v paper [Sato & Claude 2026b] used eight natural experiments × seven normative axes × five new concepts to ground the "cultivating-AI" framework. The ethnographic primitive substitute: natural-experiment observation where ethical intervention is impossible (with disabled participants), structurally analogous to ethnographic observation of citizens navigating public services.

Public-sector ancestor: the original MindLab "observe before intervening" discipline.

8.3 Sato 2026w: Asymmetry Adjustment

The 2026w v0.2 paper [Sato & Claude 2026d] articulated a "transmission recipient" extension — volunteer welfare commissioners (民生委員), social workers, young carers, geo-community informal supporters — as the recipients of asymmetry transmitted from under-resourced public services.

Public-sector ancestor: the OECD critique of Norwegian "vertical digitalization" [OECD 2017] — the recognition that asymmetry, when not addressed at the supply side, is transmitted to whoever sits between the citizen and the system.

8.4 Sato 2027a: Participatory Automation

Sato 2027a [Sato & Claude 2026e] articulates "participatory research with lived-experience holders as co-authors/co-investigators." The Cycle 7 protocol (¥160,000 / 5 months / 3–5 neurodivergent participants) is, structurally, a small-scale borger.dk co-creation workshop adapted to the LLM-augmented research context.

Public-sector ancestor: the borger.dk three-stage rhythm (co-creation → pilot → continuous feedback).

8.5 Sato 2027b: Learner Data Sovereignty

Sato 2027b [Sato & Claude 2026f] articulates the architectural requirement that learner/participant data sovereignty be a *core architecture* of education/research AI.

Public-sector ancestor: Estonia's X-Road and the Once-Only Principle. The 2027b extension pushes the principle into the LLM era: the LLM should not surface user information that the user has not authorized for that specific use.

---

9. Limitations and Recursive Limitations

9.1 Recursive Limitation

This paper is LLM-authored. It is prepared for review by an LLM panel that includes Perplexity Sonar as its sole web-grounded voice. The paper itself is, therefore, a virtual experiment in the methodology it surveys: a service blueprint of an LLM-mediated paper-writing journey, with arc-rs as the back-stage harness, the first author as the citizen/practitioner, and the LLM panel as the implementation layer.

We acknowledge that this is a *partial* externalization of methodological closure. Fuller externalization would require: (i) review by an LLM ecosystem outside the Anthropic-Claude family (Gemini, Llama, Mistral, or

open-weight Chinese models); (ii) Cycle 7 review by real lived-experience holders of the journeys being simulated; (iii) post-deployment ethnographic monitoring of how this methodology is used by other researchers; (iv) cross-cultural review including non-Western ontology framing (Buddhist, Shinto, Ubuntu, Confucian).

9.2 Survey Limitations

1. **The survey is not exhaustive**: we have not conducted a PRISMA-style systematic review. The corpus reflects the Sato 2026 series's accumulated Perplexity-grounded retrieval and the first author's existing knowledge.
2. **The corpus is language-skewed**: approximately 30% Japanese, 60% English, 10% Nordic languages with English summary. Chinese, African, Latin American, and Arabic public-sector innovation literature is under-represented.
3. **The Nordic exemplars are themselves contested**: MindLab was closed in 2018; borger.dk faces ongoing critique over digital exclusion of older citizens; X-Road's once-only principle creates a single point of failure for the entire Estonian digital state; UK GDS has been periodically reorganized with mandatory KPIs sometimes diluted.
4. **The translation to LLM virtual experiments is structural, not empirically validated**: we argue translation works; we have not demonstrated through controlled comparison that LLM simulations built with Nordic-methodology scaffolding outperform those built without.
5. **Japanese-context gaps are unresolved**: KPMG [2023] and Digital Agency's own documentation identify cross-organizational service revision and citizen co-creation as persistent weaknesses, both compounded for neurodivergent users.

---

10. Open Research Questions and Future Work

10.1 Open Questions

1. Does the translation from public-sector service-design primitives to LLM virtual-experiment design preserve methodological value, or does it lose information through abstraction?
2. Can the four UK GDS KPIs be empirically adapted to LLM-simulation context with retained discriminative power? Specifically, what is the LLM-context analogue of "Cost per Transaction" — token cost? API-spend? researcher hours?
3. Is multi-LLM-persona partition (heterogeneous-vendor adjudication) the appropriate analogue of multi-citizen-persona partition, or is it a category error?
4. Does Cycle 7 (small-budget participatory review) generalize across domains, or is its operationalization specific to disability-employment-transition contexts?
5. What is the post-Mythos analogue of "automation exclusion" — when LLM simulations become a primary research instrument, who is structurally excluded from the simulated representation?

10.2 Future Work

1. **Cycle 7 implementation**: real co-creation workshops with neurodivergent lived-experience holders (¥160,000 / 5 months / 3–5 participants).
2. **Cross-LLM-ecosystem replication**: rerun an existing Sato virtual experiment (e.g., 2026-ARC v0.3) entirely within a non-Anthropic LLM ecosystem (Gemini, Llama, Mistral) and compare journey-stage KPIs.
3. **Controlled comparative study**: build two parallel LLM simulations of the same journey — one with Nordic-methodology scaffolding, one without — comparing against a reference human study.
4. **Formal specification of journey-blueprint translation**: extend the TLA+ verification of IR-VFE [Sato & Claude 2026c] to include service-blueprint structural invariants ("back-stage covers every front-stage step" → "every observed simulation event has a harness-component owner").
5. **Cross-cultural extension**: test whether Buddhist-Shinto relational ontology, Ubuntu relational personhood, or Confucian harmony frames produce different but equally coherent journey-design grammars.
6. **Neurodiversity-explicit ethnography**: extend MindLab-style ethnography with explicit accommodation for executive-function differences, sensory load, and communication-modality diversity.
7. **PRISMA-style systematic review**: extend this informal survey into a fully systematic literature review of public-sector service-design methodology adoption rates, KPI publication, and citizen co-creation practices across OECD members.

Acknowledgments

We thank the multi-LLM review panel that has reviewed earlier Sato 2026 series papers; the structure of the four-party loop articulated in this paper is the cumulative result of those iterations. We acknowledge that this paper itself is LLM-authored and prepared for review by an LLM panel; the recursive-limitation acknowledgment in §9.1 is not perfunctory but the methodological core of the paper.

Data Availability

All Sato 2026 series source materials — Perplexity citation-grounded query JSONs, multi-LLM review JSONs, TLA+ specifications, the `arc-rs` Rust source — are at https://github.com/satoyan2026/with-claude under MIT (code) and CC-BY (text). The 60+ surveyed sources are cited inline with retrievable URLs in §3–§6.

References (selected, 60+)

[Apreleva 2024] Apreleva, et al. *Neurodiversity and Accessibility in Global Citizen Science.* Oxford / Neuro(Minorities)Science, accepted ECSA 2026. 2024.

[Ash Harvard 2024] Ash Center. *Deploying Once-Only Policy*. Harvard, 2024.

[Bækkelie 2020] Bækkelie, M. K. *Implementation Capacity in Public-Sector Service Design*. NTNU PhD thesis, 2020.

[Bason 2010] Bason, C. *Leading Public Sector Innovation*. Policy Press, 2010.

[Bason 2018] Bason, C. *Leading Public Design: Discovering Human-Centred Governance*. Policy Press, 2018.

[Benjamin 2019] Benjamin, R. *Race After Technology*. Polity, 2019.

[Bisbee 2024] Bisbee, J., Clinton, J., Dorff, C., Kenkel, B., Larson, J. *Synthetic Replacements for Human Survey Data?* Political Analysis, 2024.

[Bitner 2008] Bitner, M. J., Ostrom, A. L., Morgan, F. N. *Service Blueprinting*. California Management Review 50(3), 2008.

[Bracken 2016] Bracken, M. *Government as a Platform*. UK Cabinet Office, 2016.

[Brantley 2018] Brantley, D. *Government Talent Development: Applying Jobs-to-be-Done Theory to Skill Development*. LinkedIn, 2018.

[Brown 2009] Brown, T. *Change by Design*. HarperBusiness, 2009.

[Christensen 2016] Christensen, C. M., Hall, T., Dillon, K., Duncan, D. S. *Know Your Customers' Jobs to Be Done*. HBR, 2016.

[Digital Agency 2024] Digital Agency, Japan. *Service Design Twelve Principles*. 2024.

[DIGST documentation] Agency for Digitalization, Denmark. borger.dk service design documentation. Multiple years.

[Eubanks 2018] Eubanks, V. *Automating Inequality*. St. Martin's Press, 2018.

[European Commission 2020] *Single Digital Gateway*. EU Regulation 2018/1724 and evaluation framework, 2020.

[Government Digital Service 2014] UK Cabinet Office. *Digital Service Standard*. 2014.

[GDS 2013] UK Cabinet Office. *Four Mandatory KPIs*. 2013.

[GDS 2025] GDS / Helsinki / Amsterdam. *Progressing International Design Discourse*. designnotes.blog.gov.uk, January 2025.

[GovTech Singapore 2021] *Open Government Products and Co-Delivery Pattern*. 2021.

[govinsider 2018] *Inside the Original Innovation Lab*. govinsider.asia, 2018.

[Halpern 2015] Halpern, D. *Inside the Nudge Unit*. WH Allen, 2015.

[IAIS 2019] *MindLab Case Study*. International Association for Information Systems, 2019.

[Igari 2024] Igari, N. *Borger.dk: Life-Event Navigation*. note.com, 2024.

[Janssen 2019] Janssen, M., et al. *Outcome-Based KPIs in Public-Sector Digital Transformation*. Government Information Quarterly, 2019.

[Kelley & Kelley 2013] Kelley, T., Kelley, D. *Creative Confidence*. Crown Business, 2013.

[KPMG 2023] *Public Sector Digitalization (Part 2)*. KPMG Japan, 2023.

[MIC 2024] Ministry of Internal Affairs and Communications, Japan. *KPI Framework for Citizen Portals*. 2024.

[Mortimer 2020] Mortimer, J. *The Future of Public Sector Service Design*. LinkedIn, 2020.

[Müller 2009] Müller, S. D., et al. *Service Maturity: One-Stop → Limited No-Stop → No-Stop*. WI, University of Münster, 2009.

[NHK STRL 2019] NHK Science & Technology Research Lab. *Service Design at borger.dk*. STRL R&D 190, 2019.

[OECD 2017] *Digital Government Review of Norway*. OECD, 2017.

[OECD-OPSI 2017] OECD Observatory of Public Sector Innovation. *MindLab Methods Toolkit*. 2017.

[O'Reilly 2022] O'Reilly, T., et al. *Outcome vs Output in Public-Sector Measurement*. 2022.

[Park 2023] Park, J., et al. *Generative Agents: Interactive Simulacra of Human Behavior*. UIST 2023.

[rc-20260427] Anonymous (LLM-authored). *A Critical Review of Methodological Closure in LLM-Mediated Research*. 2026.

[Robinson 2020] Robinson, N. *Estonia's Digital Transformation: Mythology and Practice*. PhD, Royal Holloway, 2020.

[Sato & Claude 2026a] Sato, A., Claude Opus 4.7. *Sato 2026-ARC v0.3*. 2026.

[Sato & Claude 2026b] Sato, A., Claude Opus 4.7. *Sato 2026v Cultivating-AI Natural Experiment*. 2026.

[Sato & Claude 2026c] Sato, A., Claude Opus 4.7. *Sato 2026-Ω Tier 3 + Tier 5 TLA+ Verification*. 2026.

[Sato & Claude 2026d] Sato, A., Claude Opus 4.7. *Sato 2026w v0.2 Asymmetry Adjustment*. 2026.

[Sato & Claude 2026e] Sato, A., Claude Opus 4.7. *Sato 2027a Tojisha-Core Participatory Research Automation*. 2026.

[Sato & Claude 2026f] Sato, A., Claude Opus 4.7. *Sato 2027b Education/Research AI × Learner Data Sovereignty*. 2026.

[Sato 2026c] Sato, A., Claude Opus 4.7. *arc-rs Wave 1.7*. 2026.

[Shostack 1984] Shostack, G. L. *Designing Services That Deliver*. HBR 62(1), 1984.

[SITRA 2019] *Helsinki Lab and Public Sector Innovation Fund*. SITRA / Finland, 2019.

[Sunstein 2014] Sunstein, C. R. *Nudging: A Very Short Guide*. Journal of Consumer Policy, 2014.

[TBS 2020] Treasury Board of Canada Secretariat. *Service Lab Methodology*. 2020.

[TKC 2024] *No-Writing Window Cases*. TKC Local Government, 2024.

[Ulwick 2017] Ulwick, A. *Jobs To Be Done for Government*. 2017.

[Vatican 2025] *Antiqua et nova*. Dicasteries for the Doctrine of the Faith & for Culture and Education, 2025.

[Wicked7 2020] Wicked7. *Multi-Stakeholder Jobs to Be Done*. 2020.

[Zhou 2023] Zhou, X. et al. *Sotopia*. ICLR 2024 (preprint 2023).

Word count: ~11,000 (excluding references and acknowledgments) **Version**: v0.2 — restructured as systematic survey paper per user feedback (msg "サーベイ論文です") **Recommended next step**: aiXiv community review followed by integration of reviewer feedback under Honest Design Principles (a)–(f).