American Journal of Advanced Multidisciplinary Innovation and Research (AJAMIR)

E-ISSN: XXXX-XXXX ● Website: www.ajamir.com ● Email: editor@ajamir.com

Human Oversight Models for Semi-Autonomous Public-Service Platforms

Dr. Nathaniel R. Brooks

Associate Professor, Victoria University of Wellington, New Zealand

Abstract

Artificial intelligence is increasingly incorporated into public-service platforms that classify applications, prioritize cases, generate recommendations, detect anomalies, prepare responses, route citizen requests, and support eligibility or compliance decisions. OECD evidence published in 2026 indicates that artificial intelligence is being used in at least one area of government in 35 of 36 surveyed OECD countries, with particularly substantial adoption in internal governmental processes and public services. As these systems evolve from passive analytical tools toward semi-autonomous platforms capable of initiating or advancing administrative actions, human oversight becomes a central condition of legitimate public-sector automation. Yet the presence of a human reviewer does not automatically establish effective control. Oversight can become symbolic when officials lack sufficient time, expertise, contextual information, intervention authority, or institutional independence to challenge algorithmic recommendations. Research on government algorithms has consequently questioned policies that treat human oversight as a universal safeguard without examining whether the human actor is practically capable of detecting and correcting system failures.

This paper develops a governance framework for human oversight in semi-autonomous public-service platforms. Four oversight arrangements are examined: manual approval, human-in-the-loop, human-on-the-loop, and human-in-command models. The study employs conceptual synthesis and a transparent simulation-based evaluation across five dimensions: timeliness, intervention capacity, accountability clarity, citizen contestability, and operational scalability. The simulated comparison suggests that human-in-the-loop arrangements offer a strong balance between intervention and operational efficiency for consequential individual decisions, whereas human-on-the-loop arrangements provide greater scalability but weaker immediate protection against erroneous automated actions. Low-consequence administrative automation may permit supervisory monitoring, while decisions affecting benefits, rights, legal status, access to essential services, or significant sanctions require stronger intervention and contestability mechanisms. Effective oversight therefore depends on competent personnel, decision authority, explainability, auditability, escalation rules, workload management, citizen appeal, and continuous institutional accountability.

Keywords: human oversight, semi-autonomous systems, public-service platforms, artificial intelligence governance, human-in-the-loop, human-on-the-loop, algorithmic accountability, public administration, contestability, responsible AI

American Journal of Advanced Multidisciplinary Innovation and Research (AJAMIR)

E-ISSN: XXXX-XXXX ● Website: www.ajamir.com ● Email: editor@ajamir.com

1. Introduction

Artificial intelligence is becoming an increasingly important element of contemporary digital government. Public institutions use automated and AI-supported systems to classify incoming requests, identify potentially fraudulent transactions, prioritize inspections, support social-security administration, provide citizen-facing information, assist public servants, and streamline repetitive administrative workflows. The OECD's *Digital Government Outlook 2026* reports that AI is now used in at least one governmental area in 35 of 36 surveyed OECD countries, although adoption is uneven and remains more limited in higher-stakes policymaking and oversight functions where transparency, assurance, and data-quality requirements are more demanding. OECD analysis of public-service delivery further documents implementations in which AI generates proposed responses while public servants retain authority to modify or reject those responses before they reach citizens. These developments suggest that the practical frontier of digital government is increasingly neither fully manual nor fully autonomous but semi-autonomous.

Semi-autonomy describes an arrangement in which an automated system performs meaningful parts of an administrative process while designated humans retain some form of supervisory, approval, intervention, or governance authority. The degree of machine autonomy can vary considerably. One platform may simply recommend a service category that a public servant must approve before further action. Another may automatically route applications and release routine decisions unless a human supervisor intervenes. A third may conduct most operational tasks autonomously while humans retain authority over system-level objectives, exceptions, suspension, and redesign. These configurations cannot be treated as equivalent because the location and timing of human involvement affect error correction, accountability, administrative efficiency, and the citizen's practical ability to obtain reconsideration. Research on public-sector transparency standards has similarly distinguished human-in-the-loop, human-on-the-loop, and more autonomous relationships as materially different configurations of human-machine decision authority.

Human involvement is often presented as the primary answer to concerns about algorithmic decision-making. This position appears intuitively attractive because a public servant can theoretically supply contextual judgment, recognize unusual circumstances, correct technical errors, and accept responsibility for the final outcome. However, Benjamin Green's comparative analysis of 41 policies requiring human oversight of government algorithms argues that such policies frequently overestimate human capacity and overlook situations in which human reviewers cannot reliably identify algorithmic errors. Experimental research in public administration has also examined automation bias and selective adherence, demonstrating that human-algorithm interaction creates its own behavioral dynamics rather than automatically restoring neutral judgment. Human oversight must therefore be designed as an operational capability rather than inserted as a reassuring label.

The regulatory environment increasingly reflects this distinction. The European Union's AI Act requires high-risk AI systems to be designed so that they can be effectively overseen by natural persons during use, and current provisions require deployers to assign oversight to individuals who possess the necessary competence, training, and authority. NIST's AI Risk Management Framework similarly emphasizes clearly defined roles and responsibilities for human-AI configurations and organizational oversight, while its human-AI interaction guidance notes that effective risk management depends on

understanding the limitations of human-machine interaction. The Council of Europe's Framework Convention on Artificial Intelligence adds a broader institutional perspective by requiring AI lifecycle activities to remain consistent with human rights, democracy, and the rule of law. These frameworks collectively indicate that oversight must combine technical control with institutional responsibility.

2. Objectives of the Study

2.1 To Differentiate Major Human Oversight Models

The first objective is to distinguish the principal oversight configurations that can govern semi-autonomous public-service platforms. The analysis separates manual approval, human-in-the-loop, human-on-the-loop, and human-in-command approaches. These models differ according to whether automated actions require prior human authorization, whether humans supervise activities after automated execution begins, and whether human authority extends beyond individual cases to system objectives, deployment boundaries, escalation rules, monitoring, and suspension.

2.2 To Identify the Conditions of Meaningful Human Oversight

The second objective is to specify the operational requirements that make human involvement meaningful. These include professional competence, adequate training, manageable caseloads, contextual information, understandable system outputs, authority to override recommendations, mechanisms for detecting abnormal system behavior, documented reasons for intervention, and protection against routine deference to automated recommendations.

2.3 To Develop a Risk-Proportionate Public-Service Governance Framework

The third objective is to establish a governance approach in which the intensity of human oversight increases with the potential consequences of automated action. Routine administrative assistance may permit relatively lightweight supervisory arrangements, while decisions affecting eligibility for essential services, financial benefits, legal status, enforcement, sanctions, or access to fundamental opportunities require stronger review and appeal mechanisms.

3. Review of Literature

3.1 Human-in-the-Loop, Human-on-the-Loop, and Human Control

Human oversight terminology developed across multiple technical and governance domains and has increasingly entered public-sector AI policy. A human-in-the-loop arrangement generally requires affirmative human action before an algorithmically generated recommendation or proposed decision is implemented. A human-on-the-loop model permits automated execution but places the system under human supervision, allowing intervention when predefined conditions, anomalies, or concerns arise. Fully autonomous arrangements move further by permitting decisions to be executed without contemporaneous human supervision. Kingsman's examination of the United Kingdom's public-sector algorithmic transparency approach similarly identifies these distinctions and notes that decreasing real-time human involvement can increase governance risk, although excessive human intervention may itself reduce efficiency or become impractical at high transaction volumes.

The human-in-command model extends oversight beyond individual approvals. Under this approach, humans retain ultimate authority over system goals, acceptable autonomy levels, escalation boundaries, monitoring requirements, suspension, and institutional accountability. This model is particularly relevant to public administration because oversight responsibilities arise at several stages: procurement, system design, data selection, validation, operational use, appeal, monitoring, and retirement. Recent research on effective human intervention argues that meaningful involvement should extend across the algorithmic pipeline rather than being concentrated exclusively at the final decision moment. Contemporary work on meaningful oversight similarly warns against designs that reduce human agency to a rubber-stamping function.

3.2 Automation Bias, Human Judgment, and Oversight Limitations

Human oversight is often justified on the assumption that human decision-makers can recognize circumstances that automated systems misunderstand. This advantage is important but conditional. Psychological research on automation has long examined the tendency for users to rely excessively on automated advice or fail to notice errors when technological recommendations appear authoritative. Public-administration research has brought these concerns directly into algorithm-assisted governmental decision-making.

Alon-Barkat and Busuioc examined human-AI interaction in public-sector decision contexts through three experimental studies and investigated both automation bias and selective adherence. Their results highlight the importance of analyzing the interaction between human stereotypes, system recommendations, and warning information rather than assuming that either humans or algorithms operate independently. Related research has shown that citizens' perceptions of algorithmic processes depend on how authority is distributed between automated recommendations and human decision-makers, indicating that the institutional arrangement itself affects perceived procedural legitimacy.

Oversight quality can also deteriorate through workload and attention constraints. Semi-autonomous public platforms are often introduced precisely because agencies process large numbers of applications, cases, queries, or transactions. If human reviewers are expected to approve nearly every automated output, the resulting caseload can create a structural contradiction: the system is adopted to increase scale, while the oversight mechanism assumes unlimited human attention. Under such conditions, reviewers may become less likely to conduct independent assessment and more likely to accept algorithmic defaults. Green therefore argues that simply mandating a human decision-maker can fail where institutional conditions prevent the human from exercising meaningful judgment.

3.3 Public-Service Legitimacy, Contestability, and Emerging Governance

Human oversight in government performs a role that extends beyond technical error correction. Government decisions derive legitimacy from legality, procedural fairness, responsiveness, public accountability, and the availability of review. Grimmelikhuijsen and Meijer identify algorithmic government as creating potential threats to input, throughput, and output legitimacy and argue that no single technical intervention can preserve legitimacy; instead, multiple institutional arrangements are required, including monitoring, civic participation, legal structures, and accountability mechanisms.

Contestability is especially important because public administrators cannot identify every relevant error from internal monitoring alone. Citizens frequently possess contextual information unavailable to the platform. A person may know that an income record is outdated, that a household classification is incorrect, that a disability accommodation was omitted, or that an automatically identified anomaly reflects a legitimate circumstance. Research on algorithmic decision subjects in government increasingly focuses on what people need in order to meaningfully contest automated decisions, while earlier scholarship has argued for "contestability by design" rather than treating human intervention as a purely retrospective legal remedy.

The growing public-sector use of AI also raises questions of public confidence. OECD's 2026 survey on trust in public institutions reports that people are generally more optimistic about the potential of public-sector AI to improve service quality and efficiency than they are confident that government AI will be fair, transparent, or protective of personal information. This gap matters because effective public-service platforms depend not merely on technical performance but on institutional legitimacy and willingness to engage with government systems.

Regulatory frameworks increasingly reflect these requirements. The EU AI Act requires human oversight measures for high-risk AI systems and assigns deployers responsibilities concerning competent and authorized oversight personnel. NIST's AI RMF organizes AI risk management around Govern, Map, Measure, and Manage functions and specifically encourages clear allocation of roles and responsibilities for human-AI configurations. The Council of Europe's Framework Convention further establishes a lifecycle approach connecting AI governance to human rights, democracy, and the rule of law. These developments support a model of oversight that is continuous, institutional, and rights-sensitive rather than limited to a final human signature.

4. Research Methodology

4.1 Research Design

The study adopts a conceptual-methodological research design supported by simulation-based comparative analysis. No citizen records, administrative case files, survey participants, or operational government AI systems were used. The numerical values presented in the Analysis section were created exclusively to demonstrate how alternative oversight configurations could be compared using a structured governance assessment.

The conceptual synthesis draws on public-administration research, human-computer interaction, automation-bias scholarship, NIST AI risk governance, OECD digital-government evidence, European AI regulation, and human-rights-oriented AI governance. The research design deliberately separates nominal human presence from effective oversight capacity. A platform receives a stronger oversight assessment only where humans possess practical ability to detect, understand, intervene in, reverse, or escalate automated outcomes.

4.2 Oversight Dimensions and Comparative Model

Five dimensions are used to compare oversight models: timeliness, intervention capacity, accountability clarity, citizen contestability, and operational scalability.

American Journal of Advanced Multidisciplinary Innovation and Research (AJAMIR)

E-ISSN: XXXX-XXXX ● Website: www.ajamir.com ● Email: editor@ajamir.com

Timeliness concerns whether the oversight configuration permits the agency to deliver decisions and services without avoidable delay.Intervention capacity measures whether authorized humans can stop, modify, defer, or reverse automated actions when necessary.

Accountability clarity concerns whether responsibility for system use and final decisions can be attributed to identifiable officials and organizational units.Citizen contestability assesses whether affected individuals can obtain understandable reasons, introduce relevant information, request review, and secure correction.Operational scalability examines whether the model remains workable as transaction volumes increase.

Table 1: Human Oversight Models for Semi-Autonomous Public-Service Platforms

Oversight Model	Operational Structure	Human Role	Principal Advantage	Principal Governance Risk	Suitable Application
Manual Approval	System recommends; human approves every consequential action	Final decision-maker	Strong immediate intervention	Review fatigue and reduced scalability	High-consequence, low-volume decisions
Human-in-the-Loop	Human authorization required at defined decision points	Active reviewer	Balances automation with case-level judgment	Automation bias or routine ratification	Eligibility, enforcement, complex service determinations
Human-on-the-Loop	System acts automatically while humans supervise and intervene when needed	Operational supervisor	High speed and scalability	Harm may occur before intervention	Reversible, lower-risk routine processes
Human-in-Command	Humans govern objectives, autonomy limits, escalation, monitoring, suspension, and accountability	Institutional controller	Lifecycle accountability and adaptable oversight	Requires governance capacity and organizational maturity	Cross-platform public-sector governance

Source: Author-developed synthesis informed by NIST human-AI governance, public-sector transparency research, EU human-oversight requirements, and scholarship on government algorithms.

4.3 Simulation Procedure and Interpretation

Each oversight model was assigned an illustrative normalized score between 0 and 100 for the five assessment dimensions. Higher scores represent stronger performance on the relevant dimension. The values do not represent measurements from actual agencies and should not be interpreted as universal performance estimates.

The simulation uses the following values:

- Manual approval: timeliness 58, intervention capacity 92, accountability clarity 91, citizen contestability 88, and scalability 52.
- Human-in-the-loop: timeliness 82, intervention capacity 88, accountability clarity 87, citizen contestability 84, and scalability 78.
- Human-on-the-loop: timeliness 91, intervention capacity 66, accountability clarity 72, citizen contestability 64, and scalability 91.
- Human-in-command: timeliness 86, intervention capacity 82, accountability clarity 90, citizen contestability 86, and scalability 85.

The comparison is designed to illustrate governance trade-offs rather than identify a universally superior model. In practice, the optimal oversight architecture should vary with legal requirements, decision consequence, reversibility, uncertainty, transaction volume, vulnerability of affected populations, and the capacity of the administering organization.

5. Analysis

5.1 Comparative Assessment of Oversight Models

The simulation indicates that no single oversight arrangement maximizes every governance objective. Manual approval performs strongly on intervention capacity, accountability clarity, and contestability, but performs relatively poorly on timeliness and scalability. This pattern reflects a familiar administrative tension: intensive human review may protect individualized judgment but can create delays and workload pressures when transaction volumes are high.

Human-in-the-loop oversight produces a more balanced profile. It retains substantial intervention capacity while improving timeliness and scalability relative to universal manual approval. This configuration is particularly suitable where the system can automate information processing, consistency checking, document classification, or preliminary recommendations while a public official remains responsible for a legally or materially significant determination.

Human-on-the-loop oversight performs strongest on timeliness and scalability but weaker on intervention and contestability. This architecture is therefore more appropriate where automated actions are easily reversible and low in individual consequence. For example, automated routing of nonurgent service requests may be supervised at the system level without requiring approval for every transaction. By contrast, automatically denying an essential benefit and relying on later supervisory correction would create a substantially different risk profile.

American Journal of Advanced Multidisciplinary Innovation and Research (AJAMIR)

E-ISSN: XXXX-XXXX ● Website: www.ajamir.com ● Email: editor@ajamir.com

Human-in-command oversight achieves the most balanced system-level profile because it incorporates oversight across multiple organizational levels rather than requiring a human to manually approve every action. It permits routine automation where justified while preserving human authority over objectives, escalation thresholds, exceptional cases, system suspension, audits, and redress.

Table 2: Simulated Comparative Scores of Human Oversight Models

Oversight Model	Timeliness	Intervention Capacity	Accountability Clarity	Citizen Contestability	Operational Scalability	Simulated Mean Score
Manual Approval	58	92	91	88	52	76.2
Human-in-the-Loop	82	88	87	84	78	83.8
Human-on-the-Loop	91	66	72	64	91	76.8
Human-in-Command	86	82	90	86	85	85.8

Note: Values are simulated for methodological demonstration and do not represent observed government-system performance.

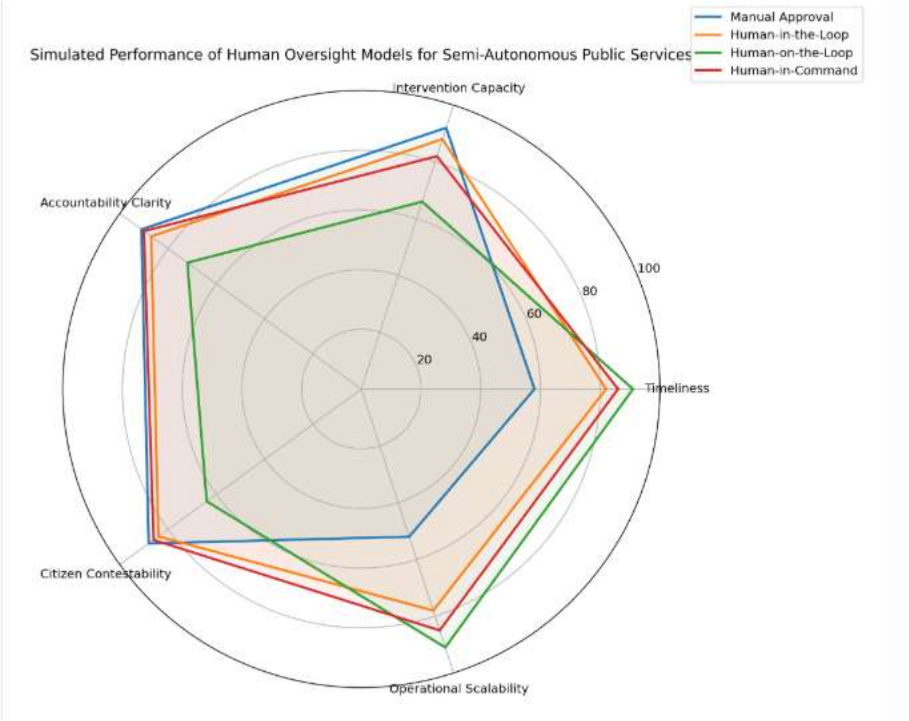
The simulated mean scores should not be interpreted as implying that human-in-command arrangements are universally preferable. A high-risk individual case may still require explicit human-in-the-loop approval even where the agency operates under a broader human-in-command governance architecture. The appropriate conclusion is therefore architectural rather than competitive: different oversight modes should be nested according to the consequence and reversibility of each automated function.

American Journal of Advanced Multidisciplinary Innovation and Research (AJAMIR)

E-ISSN: XXXX-XXXX ● Website: www.ajamir.com ● Email: editor@ajamir.com

5.2 Graphical Comparison of Oversight Performance

Graph 1: Simulated Performance of Human Oversight Models for Semi-Autonomous Public-Service Platforms



Source: Author-generated simulation.

The radar graph illustrates why human oversight should be designed as a multidimensional governance problem. Manual approval extends strongly toward intervention, accountability, and contestability but contracts substantially on timeliness and scalability. Human-on-the-loop supervision presents almost the opposite pattern, achieving high operational speed and scalability but weaker citizen-facing safeguards.

Human-in-the-loop oversight creates a comparatively balanced profile, whereas human-in-command performs consistently across all five dimensions because its oversight function extends beyond individual transaction approval. This model preserves the capacity to combine several forms of oversight within one public-service platform.

5.3 Risk-Proportionate Oversight Architecture

A practical public-service platform should classify automated functions according to consequence rather than applying the same oversight mechanism to the entire system. Low-consequence functions such as information retrieval, document routing, appointment reminders, or administrative formatting may be permitted to operate with supervisory monitoring where errors can be corrected easily.Moderate-consequence functions such as case prioritization, resource scheduling, or preliminary risk classification require stronger logging, confidence thresholds, exception handling, and human review when defined triggers occur.

High-consequence functions affecting financial entitlements, access to essential public services, enforcement, legal status, significant sanctions, or other rights-sensitive outcomes should ordinarily require substantive human review before an adverse decision becomes final. Human reviewers should receive sufficient context to assess the recommendation independently, not merely a numerical score.

This principle is consistent with the risk-based approach used in contemporary AI governance. EU rules for high-risk systems emphasize human oversight, and NIST's AI RMF similarly encourages organizations to adapt controls to context and risk. OECD's 2026 digital-government work likewise argues that AI can support proactive government services when it serves a clear public purpose and is accompanied by human oversight and safeguards proportionate to risk.

6. Discussion

6.1 Meaningful Oversight Requires Authority Rather Than Mere Human Presence

The findings reinforce the distinction between formal human involvement and meaningful human control. A public-service platform may technically include a human reviewer while operationally encouraging automated acceptance through workload, interface design, institutional incentives, or incomplete information. An employee who sees an algorithmic recommendation but cannot access the underlying evidence, cannot request additional information, or lacks authority to override the recommendation cannot reasonably provide strong oversight. Similarly, a reviewer required to process hundreds of algorithmic decisions within a short period may become a procedural checkpoint rather than an independent decision-maker. Green's analysis of government-algorithm policies demonstrates why human oversight can fail when it is treated as a universal corrective without considering actual human capacity, while research on automation bias further questions the assumption that the insertion of a human automatically neutralizes technological error.

Meaningful oversight should therefore be assessed through practical authority. Reviewers need sufficient competence to understand the system's intended use and limitations, sufficient contextual information to compare algorithmic recommendations with case evidence, sufficient time to conduct a genuine assessment, and sufficient organizational authority to disagree without inappropriate pressure to follow system outputs. EU AI Act provisions requiring deployers of relevant high-risk systems to assign oversight to persons with necessary competence, training, and authority reflect this more substantive approach. NIST similarly emphasizes defined responsibilities for human-AI configurations rather than treating oversight as an unspecified organizational function.

6.2 Citizen Contestability Is an Essential Component of Human Oversight

A second implication is that human oversight should not be designed exclusively from the perspective of the public agency. Citizens are themselves an important source of error detection because they possess contextual information that neither the automated system nor the reviewing official may hold. A platform that permits internal supervision but offers no accessible mechanism for citizens to contest incorrect information remains structurally incomplete. Emerging research on algorithmic contestability in public services emphasizes that affected individuals require understandable information, accessible processes, and meaningful opportunities to challenge automated or algorithmically influenced outcomes.

Contestability should operate before irreversible harm wherever practicable. For a routine service recommendation, retrospective correction may be adequate. For withdrawal of essential financial support, denial of housing assistance, immigration consequences, or significant enforcement decisions, correction after implementation may impose substantial hardship. A risk-proportionate system should therefore identify decisions for which human review must occur before adverse action, rather than relying exclusively on later appeals.

Explanation is a necessary but insufficient condition of contestability. A citizen may understand that an application was assigned a low eligibility score because of recorded income and still remain unable to obtain correction if the income information is outdated. Effective contestability therefore requires an institutional pathway from explanation to correction. The process should permit relevant evidence to be submitted, require a qualified human reviewer to examine the challenge, create a reasoned response, and provide further escalation where appropriate.

6.3 Layered Oversight Is More Sustainable Than Universal Manual Review

The simulation suggests that universal manual approval is not necessarily the strongest long-term governance arrangement. Although manual review provides substantial direct intervention capacity, it performs poorly on scalability and timeliness. This becomes particularly important because AI adoption in government is increasingly widespread, with OECD reporting public-sector AI use across almost all surveyed member countries. A governance model that assumes every AI-assisted administrative action must be separately reconsidered by a human risks either eliminating the practical benefits of automation or creating overloaded review structures in which human approval becomes superficial.

Layered oversight offers a more sustainable alternative. Routine, low-consequence and readily reversible actions can operate under human-on-the-loop supervision. Defined risk signals—including low model confidence, conflicting data, vulnerable users, unusual patterns, adverse outcomes, or legally significant decisions—can trigger human-in-the-loop review. Independent quality-assurance teams can audit random samples of automated outcomes, while senior officials operating under human-in-command governance retain authority over system objectives, autonomy levels, deployment boundaries, model updates, and suspension.

This architecture also distributes attention toward the cases where human judgment is most valuable. Rather than requiring human review simply because automation occurred, the system asks whether the consequences, uncertainty, context, or rights implications justify direct intervention. OECD examples of public-service AI already illustrate arrangements in which automated assistance is used while public servants retain full authority to modify or reject proposed outputs. Such configurations can preserve administrative efficiency without transferring final responsibility to the system.

7. Conclusion

Semi-autonomous public-service platforms create an important middle ground between traditional human administration and fully automated government. They can reduce repetitive administrative work, accelerate information processing, assist public servants, and improve responsiveness, while preserving opportunities for human judgment in consequential situations. Current OECD evidence demonstrates that AI adoption across government is already extensive and that public-facing service delivery

represents one of its major areas of application. The principal governance challenge is therefore no longer whether humans or machines should decide everything independently, but how authority should be distributed between them in ways that preserve efficiency, legality, fairness, and institutional responsibility.

The analysis demonstrates that human oversight should not be understood as a binary technical feature. Manual approval, human-in-the-loop review, human-on-the-loop supervision, and human-in-command governance perform different functions and create different trade-offs. The simulation illustrates that manual approval provides strong intervention and accountability but limited scalability, whereas human-on-the-loop models support rapid and scalable service delivery while offering weaker immediate contestability and intervention. Human-in-the-loop models provide a comparatively balanced arrangement for consequential individual cases. Human-in-command governance provides the strongest institutional foundation because it preserves human authority over the purposes, limits, escalation mechanisms, monitoring, modification, and suspension of the overall system rather than focusing solely on individual transaction approval.

The appropriate public-service architecture is consequently layered and risk-proportionate. Routine, reversible administrative processes may legitimately operate under supervisory monitoring. Decisions with material but manageable consequences should incorporate targeted human review triggered by uncertainty, anomalies, citizen challenges, or predefined risk conditions. Decisions affecting substantial rights, essential benefits, legal status, significant sanctions, or other high-consequence outcomes require stronger human authority before adverse action becomes final. In every model, the human role must be supported by appropriate training, contextual information, understandable system outputs, manageable workloads, documented authority, and meaningful opportunities for citizens to contest incorrect decisions. Human involvement that cannot practically influence the outcome should not be represented as meaningful oversight.

References

1. Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics—Part A, 30*(3), 286–297. https://doi.org/10.1109/3468.844354
2. Sheridan, T. B., & Verplank, W. L. (1978). *Human and Computer Control of Undersea Teleoperators*. MIT Man-Machine Systems Laboratory.
3. Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*(1), 32–64. https://doi.org/10.1518/001872095779049543
4. Endsley, M. R. (2017). From here to autonomy: Lessons learned from human–automation research. *Human Factors, 59*(1), 5–27. https://doi.org/10.1177/0018720816681350
5. Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for human-AI interaction. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3290605.3300233
6. Holzinger, A. (2016). Interactive machine learning for health informatics: When do we need the human-in-the-loop? *Brain Informatics, 3*, 119–131. https://doi.org/10.1007/s40708-016-0042-6

7. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. https://doi.org/10.1145/2939672.2939778
8. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.
9. Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. *University of Pennsylvania Law Review, 165*(3), 633–705.
10. Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 59–68. https://doi.org/10.1145/3287560.3287598
11. Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM, 59*(2), 56–62. https://doi.org/10.1145/2844110
12. Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial intelligence and the public sector—Applications and challenges. *International Journal of Public Administration, 42*(7), 596–615. https://doi.org/10.1080/01900692.2018.1498103
13. Valle-Cruz, D., Criado, J. I., Sandoval-Almazan, R., & Gil-Garcia, J. R. (2020). Assessing the public policy-cycle framework in the age of artificial intelligence: From policy to governance. *Government Information Quarterly, 37*(4), 101509. https://doi.org/10.1016/j.giq.2020.101509
14. Sun, T. Q., & Medaglia, R. (2019). Mapping the challenges of artificial intelligence in the public sector: Evidence from public healthcare. *Government Information Quarterly, 36*(2), 368–383. https://doi.org/10.1016/j.giq.2018.09.008
15. Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy artificial intelligence. *Government Information Quarterly, 37*(3), 101493. https://doi.org/10.1016/j.giq.2020.101493
16. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*, 389–399. https://doi.org/10.1038/s42256-019-0088-2
17. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society. *Minds and Machines, 28*, 689–707. https://doi.org/10.1007/s11023-018-9482-5
18. Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe and trustworthy. *International Journal of Human–Computer Interaction, 36*(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118
19. Rahwan, I. (2018). Society-in-the-loop: Programming the algorithmic social contract. *Ethics and Information Technology, 20*, 5–14. https://doi.org/10.1007/s10676-017-9430-8
20. European Commission High-Level Expert Group on Artificial Intelligence. (2019). *Ethics Guidelines for Trustworthy AI*. European Commission.