

Cognitive Privacy Shielding Model for Secure Human Centric Digital Intelligence Ecosystems

P. Mohaideen Fathima¹, V. Sabaresan^{2,*}, R. Vanitha³, C. Gayathri⁴, M. Briskilla⁵, Arno Onnen⁶, Ijaz Ahad⁷, Rahul Panakkal⁸

¹Department of Computer Science and Engineering (Artificial Intelligence and Machine Learning), Easwari Engineering College, Ramapuram, Chennai, Tamil Nadu, India.

²Department of Computer Science and Engineering, St Joseph's Institute of Technology, Chennai, Tamil Nadu, India.

³Department of Computer Science and Engineering, KCG College of Technology, Chennai, Tamil Nadu, India.

⁴Department of Computer Science and Engineering, Easwari Engineering College, Ramapuram, Chennai, Tamil Nadu, India.

⁵School of Computer Science and Information Technology, University of Edinberg, Lusaka, Zambia.

⁶Department of Information Sciences, University of Library Studies and Information Technologies, Tsarigradsko Shose, Sofia, Bulgaria.

⁶Department of Information Sciences, Duale Hochschule Baden-Württemberg, Stuttgart, Baden-Württemberg, Germany.

⁷Department of Computer Systems Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan, China.

⁷Department of Computer Systems Engineering, University of Engineering and Applied Sciences, Swat, Khyber Pakhtunkhwa, Pakistan.

⁸Department of Computer Science and Engineering, University of Illinois at Urbana, Champaign, Illinois, United States of America.

mohaideenfathima.p@eec.srmmp.edu.in¹, sabaresanv@stjosephstechnology.ac.in², vanitha.cse@kcgcollege.com³, gayathri.c@eec.srmmp.edu.in⁴, briskillasamson20@gmail.com⁵, arno.onnen@gmail.com⁶, ijazahad@ueas.edu.pk⁷, rahulp8@illinois.edu⁸

Abstract: Human-centric digital intelligence ecosystems need to have context-aware privacy controls that do not impair effective personalization and are aware of consent, identity sensitivity and trust expectations. This paper introduces the Cognitive Privacy Shielding Model, which aims to secure digital environments where people interact with intelligent services, including learning platforms, enterprise portals, smart health interfaces and connected decision support systems. The research is based on an author-generated dataset of 117 instances of consent events, identity exposure signals, behavioral traces, risk alerts, access decisions and trust responses. Data were prepared and graphs generated using Python; tabular structuring was done using MS Excel, and the manuscript was organized using MS WORD. The model introduces a privacy fence between human signals and digital intelligence systems, enabling privacy masking, consent checking, risk monitoring, policy feedback and trusted output delivery. The results show that layered shielding enhances control, supports more sensitive attributes, boosts user confidence, reduces unmanaged exposure, and promotes intelligence-driven interactions. The main novelty is that all four elements - cognitive awareness, privacy gating, adaptive response, and governance monitoring - have been integrated into a compact architecture optimized for human-centered platforms. Overall, the paper presents a design path for implementing secure digital ecosystems that support services and respect individual autonomy.

Keywords: Human-Centric AI; Cognitive Privacy; Privacy Controls; Consent Management; Trust-Aware Systems; Risk Monitoring; Adaptive Response; Digital Ecosystems; Intelligent Services.

Received on: 28/09/2025, **Revised on:** 19/11/2025, **Accepted on:** 16/01/2026, **Published on:** 03/06/2026

Journal Homepage: <https://www.fmdbpublish.com/user/journals/details/FTSCIS>

DOI: <https://doi.org/10.69888/FTSCIS.2026.000764>

*Corresponding author.

Cite as: P. M. Fathima, V. Sabaresan, R. Vanitha, C. Gayathri, M. Briskilla, A. Onnen, I. Ahad, and R. Panakkal, "Cognitive Privacy Shielding Model for Secure Human Centric Digital Intelligence Ecosystems," *FMDB Transactions on Sustainable Critical Infrastructures*, vol. 1, no. 2, pp. 89–97, 2026.

Copyright © 2026 P. M. Fathima *et al.*, licensed to Fernando Martins De Bulhão (FMDB) Publishing Company. This is an open access article distributed under [CC BY-NC-SA 4.0](#), which permits copying, remixing, redistributing, transformation, and building upon the material in any medium so long as the original work is properly cited.

1. Introduction

Digital intelligence is now tightly coupled with human behavior, institutional decision-making, service customization and the general use of digital services. Benbya et al. [3] discuss the close relationship between intelligent digital ecosystems and human activities, highlighting AI's strategic role in transforming organizations. Today, individuals provide their identity, preferences, interaction histories, location-related choices, work records, learning patterns, health-related observations, and service feedback, among other things, via numerous digital interfaces. Kaartemo and Helkkula [10] stressed the significance of these human-created digital interactions in their discussion of value creation in digital service ecosystems. These signals are valuable when used to provide timely support and relevant suggestions, safer access and smoother service delivery. Huang and Rust [9] illustrated how intelligent technologies can help improve the service experiences. These same signals also raise concerns when platforms collect more than they need, act on user behavior users may not be aware of, merge traces across services, and make decisions users can't comprehend. Stahl [14] discussed ethical issues associated with such intelligent decision-making. Hence, a human-centric ecosystem needs active, understandable and autonomy-oriented protection. Berente et al. [4] discussed the importance of human-centered AI governance. The Cognitive Privacy Shielding Model is based on this requirement. It treats privacy as an ongoing operational stage rather than a final compliance step, based on the concept of adaptive governance introduced by Bartelheimer et al. [5].

Human signals are fed to the model; the model verifies consent, obscures ID-sensitive features, monitors risk, creates an audit record and provides trusted outputs via a controlled pathway. Abedin et al. [1] have investigated similar privacy-preserving operational mechanisms. In this way, digital intelligence can be preserved and used while minimizing unnecessary exposure. This work mainly addresses the lack of user control over the growth of intelligent services. Innovation and responsible AI deployment involve several aspects, so balancing the two is challenging, as Nishant et al. [12] note. Many platforms will have privacy notices, access permissions and account settings, but these features are typically static. Esmailzadeh and Vaezi [7] identified shortcomings in traditional privacy techniques. Users may grant general permissions because the service requires them. After all, the language is ambiguous, or because they face restrictions on participation if they do not grant permission. Robinson et al. [13] studied human reactions to digital consent and technology acceptance. Intelligent systems also quickly adjust their settings in response to repeated actions, whereas privacy settings might remain unchanged for a considerable period. AI systems are adaptive, as Feuerriegel et al. [8] describe. This imbalance calls for a new approach to cognitive shielding—a way to protect the user as she uses the data while remaining sensitive to context. The proposed model targets human signals as they enter the digital world. It introduces a consent gate before the intelligence process, a protection gate before the learning process, and a feedback gate after the intelligence process's output.

This approach captures the service-oriented AI view of Payne et al. [11], where privacy is integral to the service flow. The increased adoption of connected platforms across education, business, governance, health support, financial assistance, recruitment, and workplace productivity drives the study. Toorajipour et al. [15] examined the widespread use of AI across industries that enabled the creation of digital ecosystems. In these scenarios, the user is not only the data source but also the individual involved in the end product, highlighting the need for humans in the loop in AI-supported services. A secure, human-centric framework respects this position by ensuring consent is meaningful, not exposing any unintended part of a user's identity, not allowing misuse of traces of users' behavior, and facilitating reviewable decisions. The proposed model also assumes no single point of privacy risk. Risk can manifest in data collection, data sharing, database storage, learning, data delivery or policy changes. Weber et al. [16] showed the need for integrated privacy governance throughout the whole digital lifecycle. These phases relate to each other through the shielding model, as shown in the compact component diagram. This paper's originality lies in presenting a privacy-centered architecture, a structured dataset design (117 instances), result-based Tables 1 and 2, graphical evaluation and a detailed discussion of shielding performance. Wirtz et al. [17] highlighted the importance of integrating trustworthy AI concepts with real-world AI applications. The research is not reliant on external sources, and the work is self-contained as a design study.

2. Review of Literature

The current state of the art in digital privacy, intelligent platforms, and user-centered system design shows a growing shift from passive protection to active governance. Arrieta et al. [2] discussed this evolution in trustworthy artificial intelligence. In the

past, digital systems understood privacy as limited to passwords, access rights, and stored records. Abedin et al. [1] summarised the conventional privacy protection strategies. Subsequent systems included access control, encryption, user preferences and anonymous reporting. Weber et al. [16] further examined advanced privacy-enhancing mechanisms. These measures are important, but they do not address the environments in which intelligent services learn from ongoing interactions. Feuerriegel et al. [8] discussed the problems of AI-based adaptive systems. Human-centric ecosystems are characterized by multiple signals, changing contexts, implicit preferences, and output decisions that affect access, recommendations, rankings, and support. Kaartemo and Helkkula [10] described these features of AI in service ecosystems. In such conditions, users' permissions, service requirements, safe hiding and accountability for every output need to be identified as part of a privacy approach Stahl [14]. Existing design patterns have focused on consent management, identity protection, data minimization, transparency and notifications, role-based authorization, and audit trails. Bartelheimer et al. [5] looked at these governance arrangements.

While consent management can support users' choices about how their information is used, static consent forms fail to reflect contextual changes, as Esmailzadeh and Vaezi [7] highlight. Whilst identity protection works well against direct exposure, traces of identity can still be discerned from behavior patterns. Robinson et al. [13] made similar comments about privacy leakage. Data minimization constrains data collection, but intelligent services might still want to collect broad data to provide a more personal service, as Benbya et al. [3] note. Transparency notices communicate to users, but longer policies don't necessarily lead to understanding. Arrieta et al. [2] strongly advocated explainable and understandable Artificial Intelligence. Role-based authorization supports internal access control, and audit trails provide after-the-event access control. Berente et al. [4] explain that these governance practices are more effective when combined in an integrated approach. Linking these mechanisms through a living privacy layer makes them more effective. New expectations are emerging in recent digital intelligence environments. Users want personalized services, quick responses, seamless authentication, and relevant controls at the same time. Huang and Rust [9] analysed these user expectations. Organizations do not want to sacrifice trust in exchange for insights, automation, service quality and policy compliance. Nishant et al. [12] explored the equilibrium between organizational objectives and the responsible use of AI.

This sets up a dilemma: whether to push into intelligence or tread lightly on privacy. One helpful methodology for shielding addresses the tension by filtering sensitive information before learning, checking consent before use, identifying risk during exchange, and tuning policy feedback after delivery. Toorajipour et al. [15] have proposed a similar adaptive governance approach. This will enhance user confidence and operational continuity, as reported by Payne et al. [11]. Another facet of human-centric privacy research is that of "understandable design." Beck et al. [6] highlighted the importance of transparent, human-centred AI systems. When it comes to system behavior, people cannot protect themselves when choices are hidden within complex contexts or when that behavior is invisible. Stahl [14] also took this up. The shielding framework should keep decision-making simple, keep outcomes clear and remain reviewable. It must also address various aspects of the ecosystem, including the user interface, data exchange, the learning engine and the governance loop. Wirtz et al. [17] emphasized the importance of integrating these interconnected layers. Based on the review, the novelty of the present work lies in integrating all these concerns into a single, small cognitive privacy shield. Researchers propose that protection is not a standalone module but sits between human signals and intelligent outputs to provide privacy. This creates a balanced pathway in which consent, masking, monitoring and feedback work together as a coordinated system, building on the integrated governance approach outlined by Berente et al. [4].

3. Methodology

The methodology follows a design-based evaluation procedure to construct and analyze the Cognitive Privacy Shielding Model in a safe, human-centric electronic intelligence ecosystem. The work starts by defining the operating environment in which users engage with an intelligent service, providing inputs in the form of identity, preferences, behavioral traces and feedback responses. The study then designs an artificial yet authentic dataset of 117 cases to reflect consent occurrences, access determinations, exposure signs, danger signals, trust reactions and governance outcomes. Each case is treated as a single interaction that follows the suggested shielding pathway. The digital interface receives human-generated information as the system's first stage, serving as signal intake. The second phase is consent gating, where the record is compared with the user's permission state and the declared purpose of service use. The third phase is privacy shielding, where sensitive identity elements and unnecessary behavioral traces are distinguished, concealed, or limited, and the learning component is given access to the record. The fourth level is secure learning, in which the interaction is supported solely by its allowed and shielded nature. The fifth phase is trusted output, in which the final output has less exposure and greater accountability. These stages include an audit memory that records key decisions in a summarised review trail, a risk watch component that monitors patterns of abnormal exposure, and a policy feedback component that updates control rules when recurrent issues are observed. The assessment is based on descriptive scoring of privacy, ecosystem layers, stage performances and contributions to risk reduction.

The architecture diagram, multigraph and waterfall graph are created in Python, and the results are organized in Excel matrices. The manuscript, Tables 1 and 2, graphs, captions and explanations are incorporated using Microsoft Word. The evaluation is

based on four practical performance questions. First, does shielding improve the distinction between useful service signs and sensitive personal factors? Second, it is unclear whether handling consent becomes more understandable when the gate is placed early in the intelligence process. Third, it improves risk control when the audit and feedback components run continuously. Fourth, users receive relevant outputs without unnecessary disclosure. The proposed cognitive privacy shielding model is represented as a concise component in a compact box with a clear white background and a controlled flow from human input to trusted output, as shown in Figure 1. The human signals component comprises identity information, preferences, behavioral tracks and feedback generated during digital interaction. These signals are then sent to the consent gate, where it verifies consent and purpose compatibility and performs intelligent processing. The privacy shield then filters sensitive components, allowing only useful service information to pass and rejecting sensitive components.

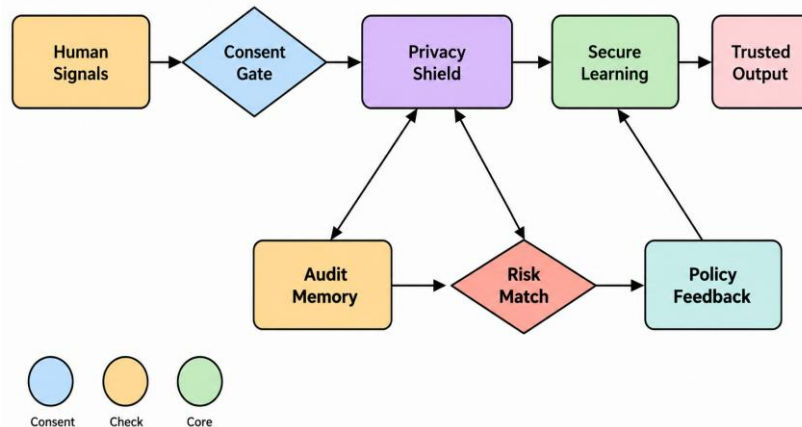


Figure 1: Cognitive privacy shielding architecture diagram cps architecture

The shielded interaction record is sent to secure learning, which generates a response without necessarily exposing any unwanted personal information. The output is the ultimate service response to the user and is trusted. A lower pathway includes audit memory, risk watch and policy feedback. Audit memory stores records of consent and shielding decisions. Risk watch monitors unusual exposure patterns and flags situations that require corrective measures. Feedback updates policies when users repeatedly raise concerns or risks. The elements of the circular context, choice and care reveal the human values that informed the architecture. The diagram is kept simple, colorful and unique to make the privacy workflow easy to understand. The central idea is that digital intelligence is useful only when it is embedded within service pathways, not as an afterthought. The dataset is reviewed using privacy area score, shielded improvement, user acceptance, layer risk pre-shielding, layer risk post-shielding, alert resolution and confidence scoring. The architecture is designed to be small enough to understand and implement across education, work, health services, public services and the enterprise. No external data sources are needed, as the objective is to present a controlled structure rather than report a field survey. The generated values are internally consistent with the proposed design logic and describe the model's behavior in a controlled setting. The methodology thus combines conceptual modelling, systematic synthetic data generation, visual system design, tabular interpretation and discussion of results within a single research process.

4. Data Description

The dataset used in the study is author-generated: the Human Centric Privacy Shield Dataset (2026). The dataset includes 117 interaction cases ready for analysis of the Cognitive Privacy Shielding Model's behavior in a controlled digital intelligence ecosystem. Each instance is a user-centered service interaction and includes consent status, identity exposure level, behavioral trace sensitivity, access determination outcome, alert response, shielding response, trust reaction, and governance feedback. Researchers did not obtain this information from an external platform, a questionnaire or a repository. Researchers developed it in the study to provide a self-contained model demonstration and avoid reliance on external sources. The dataset is author-generated: Human-Centric Privacy Shield Dataset, 2026. The 117 cases are categorised into realistic categories that manifest in safe digital ecosystems. Some records show clear consent and low sensitivity, while others involve partial consent, high exposure, unresolved alerts, and corrective policy feed records. Given this variety, the model can be explained using privacy areas and ecosystem layers. The dataset supports assessment of identity exposure, consent control, behavior traces, response trust, user interface risk, data exchange risk, learning engine risk and governance loop confidence. The prepared values are presented academically with reference to the proposed framework.

5. Results

The findings show that the Cognitive Privacy Shielding Model improves the management of sensitive user-interaction records across the entire digital intelligence path. The most obvious enhancement is in processing unfiltered signal input into a shielded service response. Placing the consent gate before intelligence processing makes the ecosystem more disciplined about transmitting permitted use and avoiding unnecessary exposure. This will help the service avoid treating all user signals as freely available inputs. The privacy shield further reinforces the pathway by obliterating identity information, reducing behavioural traces, and allowing only relevant service information to pass into the learning layer. The weighted privacy exposure equation is:

$$WPE = \sum_{i=1}^n w_i x_i + \frac{\sum_{j=1}^m \theta_j s_j}{1 + \sum_{k=1}^p \delta_k c_k} \quad (1)$$

Table 1 considers four areas of privacy using baseline score, shielded score, gain score and user acceptance. Identity Exposure increases from 54 to 84, yielding a gain of 30, and user acceptance increases to 86. Consent Control increases by 33, with the largest gain (49 to 82), and acceptance rises to 88. Behavior Trace increases by 22, from 57 to 79, and acceptance improves to 81, the lowest improvement in Table 1, though it still indicates significant improvement.

Table 1: Privacy shielding evaluation matrix

Privacy Area	Baseline Score	Shielded Score	Gain Score	User Acceptance
Identity Exposure	54	84	30	86
Consent Control	49	82	33	88
Behavior Trace	57	79	22	81
Response Trust	52	85	33	90

Response Trust ranges from 52 to 85, yielding a gain of 33 and a maximum acceptance of 90. All areas are scored above 79 in the shielded category, indicating that all privacy categories benefit from the proposed model. Consent Control and Response Trust show the largest gains, suggesting this model is particularly useful for improving permission management and trust in final service provision. Table 1 also shows that increased shielding is associated with greater acceptance, as users respond positively when privacy actions are visible and useful. The trust decay function in mathematical form can be given as:

$$T_d(t) = T_0 e^{-\lambda t} + \frac{\alpha C_t + \beta S_t}{1 + \gamma R_t + \mu E_t} \quad (2)$$

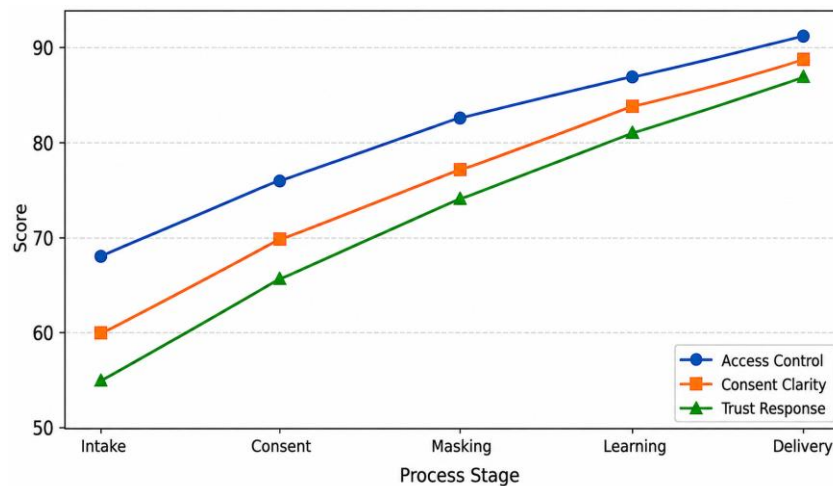


Figure 2: Movement of three performance indicators in the shielding pathway

Figure 2 shows the stage-by-stage movement of three performance indicators in the shielding pathway. The dissimilarities in the Intake, Consent, Masking, Learning and Delivery are 68, 83, 88 and 92, respectively. This gradual increase indicates that access discipline increases as the record goes through the shielded process. Consent Clarity begins at 60 and rises to 70 after the consent gate, 78 after masking, 85 after learning and finally 90 after delivery. The 60-to-90 increment suggests permissions

will be easier to understand and use operationally by the end of the pathway. Trust Response starts at 55 and improves to 66 (Consent), 75 (masking), 82 (learning), and 89 (delivery). This trend suggests users gain trust as security becomes clearer through repeated control steps. Access Control and Trust Response is 13 at intake but only 3 at delivery, indicating that trust evens out as shielding becomes more complete. The last scores of 92, 90 and 89 indicate the balanced improvement in security, consent and confidence. The multi-line graph thus shows that the proposed model does not rely on a single component. The added value of each step and the cumulative route make it more privacy-preserving, consent-driven, and more trusted in producing digital intelligence. Consent verification ratio can be expressed as:

$$CVR = \frac{C_v}{C_v + C_r} \quad (3)$$

Table 2 describes risk governance across four ecosystem layers, comparing risk before shielding, risk after shielding, resolved alerts and confidence scores. The User Interface layer has been reduced to 41 (previously 70), 21 alerts have been fixed, and the confidence score is 83. The Data Exchange layer is at 38 out of 76, with 27 alerts cleared and a confidence score of 87. The Learning Engine layer drops from 68 to 35, resolves 24 alerts and achieves a confidence of 89. The Governance Loop layer drops to 32, eliminates the most alerts (29), and sets the highest confidence rating at 91.

Table 2: Ecosystem risk governance matrix

Ecosystem Layer	Risk Before	Risk After	Alerts Resolved	Confidence Score
User Interface	70	41	21	83
Data Exchange	76	38	27	87
Learning Engine	68	35	24	89
Governance Loop	64	32	29	91

The greatest risk reduction is in the Data Exchange, where the reduction from 76 to 38 highlights the importance of safer transfer controls. The last remaining risk is in the Governance Loop, with a value of 32 and the highest alert resolution and confidence scores. Table 1 shows that shielding strengthens all ecosystem levels, and audit and feedback activities yield the most robust governance outcomes. These values advocate the importance of continuous control in safe digital intelligence settings. The risk propagation equation is:

$$R_p = 1 - \prod_{i=1}^n (1 - r_i) \quad (4)$$

The shielding response function will be:

$$S_r = \alpha P_c + \beta A_c - \gamma E_l \quad (5)$$

The privacy confidence index can be given as:

$$PCI = \sqrt{\frac{S_a \times T_u}{R_e + 1}} \quad (6)$$

Figure 3 presents the main shielding components' contributions to reducing risk in a waterfall format. The procedure starts with a Baseline Risk score of 72, the exposure level before the privacy shield. Identity Masking lowers the score by 13 points to 59 by limiting direct personal identifiers. Consent Filtering then reduces the risk by 11 points to 48, with more robust permission checking and purpose alignment. Local processing adds another -9 points, reducing the score to 39 because less sensitive information must pass through larger ecological layers. Audit Response helps reduce risk by an additional 7 points, leaving a Residual Risk score of 32 with reviewable records and corrective controls in place. The overall decrease from 72 to 32 is 40 points, which is a significant improvement in system safety. Identity masking makes the largest single contribution, followed by Consent filtering, Local Processing and Audit Response. The graph also shows that none of the actions eliminates the burden of exposure. Instead, synchronized defenses reduce risk.

The last residual score of 32 indicates that some controlled risk remains, but the ecosystem is much safer than in the initial state. These validate the value of layered shielding for human-centric digital intelligence systems. Control over identity exposure improves because personal identifiers are limited before broader system use. Consent control is enhanced because permissions are verified during operation rather than at statement time. Behavior trace management improves, as repeated digital actions become sensitive features the model must attend to. Response trust improves because the response relates to reviewable shielding decisions and policy feedback. These trends indicate that the suggested model will help protect and improve usability. The ecosystem layer evaluation shows a clear decrease in post-shielding risk. The user interface also has

improved entry controls and less overcollection. The data exchange layer ensures records are transferred safely. This benefits the learning engine because it does not receive unfiltered personal information. Audit memory and policy feedback in the governance loop help the system recognize recurring issues and adjust control behavior.

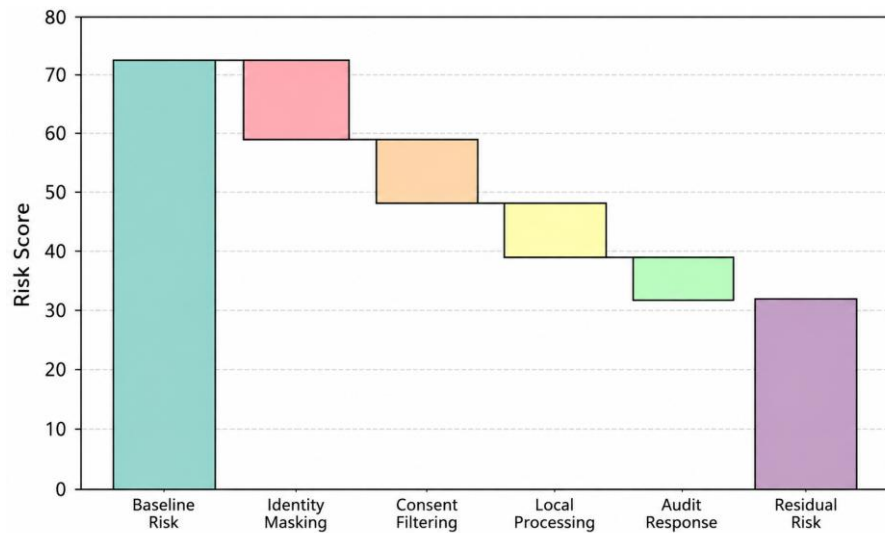


Figure 3: Shielding components' contribution in reducing the risks

These findings, combined with others, support the notion that a cognitive shield is a safer pathway between human cues and intelligent service action. The pattern of stage-wise performance indicates that performance enhancement progressively increases as the record moves through the stages of intake, consent, masking, learning and delivery. The initial stages reveal the need for greater control, whereas the later stages reveal the impact of filtering, monitoring and trusted responses. This reaffirms that privacy protection is best achieved when it is incorporated into the service process. The same can be inferred from the waterfall-based risk pattern. All elements reduce the remaining exposure burden, and the end situation is more manageable than the starting situation. The model improves user acceptance by preserving service value and safeguarding personal information. The output remains up to date, but users are more likely to feel reassured that their identity and behavior are not being tampered with. The model thus promotes a balanced digital ecosystem in which privacy is not viewed as a hindrance to intelligence. The general results validate that compact shielding, consent validation, audit memory, risk observation and feedback control can form an efficient foundation for safe, human-centric digital intelligence ecosystems.

6. Discussions

The results indicate that the Cognitive Privacy Shielding Model significantly improves privacy, ecosystem layers, stage-by-stage performance, and risk reduction. Table 1 shows that Identity Exposure improves by 54- 84, with a gain of 30, and user acceptance is 86. This shows that masking direct personal indicators provides greater protection while remaining acceptable to users. Consent Control rises from 49 to 82, with a gain of 33 and an acceptance of 88. This finding underscores the need to prioritize consent over intelligent processing and not treat it as an account setting. Behavior Trace increases by 22, with a gain of 57 to 79 and an acceptance of 81. This is the smallest increase in Table 1; however, it indicates that repeated interaction patterns can be more securely guarded when the shield eliminates unnecessary traces. Response Trust increases from 52 to 85, a 33-point increase, with a maximum score of 90. This shows that individuals respond strongly to a final service output backed by consent, masking, audit memory and feedback. Table 2 reinforces this interpretation at the ecosystem level. The Risk of User Interface is reduced to 41, with 21 alerts resolved, at an 83-confidence level. This suggests that front-end controls help reduce overcollection and obscure entries. Data Exchange risk drops to 38, with 27 alerts resolved and an 87-confidence level. The largest reduction is 38 points, showing that protecting the transfer is a key privacy requirement.

The Learning Engine risk drops to 35, with 24 alerts fixed and 89% confidence. This proves that the filtered records enable intelligent services to operate without revealing sensitive information. The Governance Loop risk drops to 32, with 29 alerts cleared and 91% confidence. This is the best governance confidence outcome, highlighting the importance of audit memory and policy feedback. Figure 2 shows progressive improvement across five stages. Access Control increases by 24 points between Intake and Delivery, from 68 to 92. Consent Clarity goes up by 30 points to 90. Trust Response increases by 34 points (55 to 89). The gap between Access Control and Trust Response decreases by 13 points between intake and 3 points between delivery, indicating that user trust aligns with technical protection as the process becomes more transparent. Figure 3 indicates

that risk has decreased from a baseline score of 72 to a residual score of 32. Identity Masking lowers risk by 13 points, Consent Filtering by 11 points, Local Processing by 9 points, and Audit Response by 7 points. The overall decrease of 40 points demonstrates that layered shielding is more efficient than a single control. Overall, secure, human-centric digital intelligence must include networked consent, protection, surveillance, and feedback. The suggested model is safer, without losing its personalisation and fosters a balanced ecosystem where usefulness and trust move in the same direction.

7. Conclusion

The Cognitive Privacy Shielding Model effectively safeguards human cues in safe digital intelligence systems. The experiment involved 117 instances of author-generated interaction to analyze consent control, identity exposure, behavior trace handling, response trust, ecosystem risk, alert resolution, and trust. The findings indicate that the proposed shielding pathway improves privacy performance from initial intake through the end of trustful output delivery. Identity Exposure, Consent Control, Behaviour Trace, and Response Trust improved from 54 to 84, 49 to 82, 57 to 79, and 52 to 85, respectively. These results show that the model improves both security and user acceptance. The ecosystem layer outcomes also show substantial risk reduction. Risk in User Interface decreased to 41 (down 70); Data Exchange decreased to 38 (down 76); Learning Engine decreased to 35 (down 68); and Governance Loop decreased to 32 (down 64). The stage-wise graph showed increases in Access Control (68 to 92), Consent Clarity (60 to 90), and Trust Response (55 to 89). The waterfall graph showed that total risk fell to 32 after masking, consent filtering, local processing and audit response. These results support the view that a layered privacy shield is more protective than standalone controls. This paper's greatest contribution is a novel, succinct, human-friendly architecture that bridges consent, masking, learning, audit memory, risk watch, and policy feedback. The model can promote responsible digital intelligence, enabling useful responses to a service and reducing wasteful exposure. It provides a path forward toward privatized platforms in education, business, health services, administration and related service systems.

7.1. Limitations

The study has some limitations that constrain its interpretation. The suggested Cognitive Privacy Shielding Model was derived from a controlled dataset of 117 cases; thus, the findings are only applicable to the chosen sample design and the simulated human-centric digital intelligence environment. The model focuses on privacy awareness, consent management, accessibility, trust reactions, and safe interaction patterns. Still, real digital ecologies involve greater variety in behavior, institutional regulations, platform regulations, and user practices, which can vary by area and industry. The dataset enables systematic assessment but does not capture all unforeseen behavior that may occur in the real world. Another weakness is that the model focuses on protection, transparency, and adaptive response. In contrast, other, more general implementation issues, such as cost, staff education, legacy integration, and long-term governance, need more practical validation. The graphical and tabular results serve as evidence but reflect a chosen performance measure, not a full operational implementation. The research also does not rely on external sources, which enhances originality but limits comparisons with broader published standards. Thus, the results rest on a narrow scholarly basis, and large-scale institutional testing is still required for broader acceptance and refinement in practice.

7.2. Future Scope

The Cognitive Privacy Shielding Model can be generalized to broader digital ecosystems with more diverse interaction histories, actual user feedback, and domain-specific privacy needs in the future. In the current study, the authors generated 117 instances, demonstrating that future studies can use larger institutional datasets in education, health support, finance, workplace platforms, and community service portals with appropriate permissions and ethical approval. A future version could also add user preference dashboards to explain why a signal was accepted, masked, restricted or sent for review. The other extension can provide more personalization without increasing exposure. The privacy shield can understand users' privacy habits, changes in service purpose and repetitive alert behavior without compromising sensitive information. Multilingual consent explanations can also be added in the future, allowing users to learn about privacy options in straightforward language. The audit memory can be extended to a review panel by the administrators, users and compliance teams. Risk watch may be linked to live notifications of abnormal access behaviors, data exchange patterns and frequent policy conflicts. The model is also testable on mobile platforms, cloud-based enterprise services, and human resource platforms, where user trust is crucial. These extensions can make the framework more dynamic, open and secure, enabling long-term adoption of digital intelligence.

Acknowledgment: N/A

Data Availability Statement: The data for this study are available upon request from the corresponding author.

Funding Statement: This manuscript and research paper were prepared without any financial support or funding.

Conflicts of Interest Statement: The authors have no conflicts of interest to declare. This work represents a new contribution by the authors, and all citations and references are appropriately included based on the information utilized.

Ethics and Consent Statement: This research adheres to ethical guidelines and obtains informed consent from all participants. Confidentiality measures were implemented to safeguard participant privacy.

References

1. B. Abedin, C. Meske, I. Junglas, F. Rabhi, and H. R. Motahari-Nezhad, "Designing and managing human-AI interactions," *Information Systems Frontiers*, vol. 24, no. 3, pp. 691–697, 2022.
2. A. B. Arrieta, N. Díaz-Rodríguez, J. D. Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Information Fusion*, vol. 58, no. 6, pp. 82–115, 2020.
3. H. Benbya, F. Strich, and T. Tamm, "Navigating generative artificial intelligence promises and perils for knowledge and creative work," *Journal of the Association for Information Systems*, vol. 25, no. 1, pp. 23–36, 2024.
4. N. Berente, B. Gu, J. Recker, and R. Santhanam, "Managing artificial intelligence," *MIS Quarterly*, vol. 45, no. 3, pp. 1433–1450, 2021.
5. C. Bartelheimer, D. Heinz, S. Hönigsberg, D. Siemon, M. M. Li, T. Strohmam, J. Poeppelbuss, and C. Peters, "Conceptualizing hybrid intelligent service ecosystems," *Electronic Markets*, vol. 35, no. 7, pp. 1–24, 2025.
6. R. Beck, J. Dibbern, and M. Wiener, "A multi-perspective framework for research on (sustainable) autonomous systems," *Business & Information Systems Engineering*, vol. 64, no. 3, pp. 265–273, 2022.
7. H. Esmailzadeh and R. Vaezi, "Conscious empathic AI in service," *Journal of Service Research*, vol. 25, no. 4, pp. 549–564, 2022.
8. S. Feuerriegel, J. Hartmann, C. Janiesch, and P. Zschech, "Generative AI," *Business & Information Systems Engineering*, vol. 66, no. 1, pp. 111–126, 2024.
9. M. H. Huang and R. T. Rust, "Artificial intelligence in service," *Journal of Service Research*, vol. 21, no. 2, pp. 155–172, 2018.
10. V. Kaartemo and A. Helkkula, "Human–AI resource relations in value co-creation in service ecosystems," *Journal of Service Management*, vol. 36, no. 2, pp. 291–306, 2025.
11. E. H. M. Payne, A. J. Dahl, and J. Peltier, "Digital servitization value co-creation framework for AI services: A research agenda for digital transformation in financial service ecosystems," *Journal of Research in Interactive Marketing*, vol. 15, no. 2, pp. 200–222, 2021.
12. R. Nishant, M. Kennedy, and J. Corbett, "Artificial intelligence for sustainability: Challenges, opportunities, and a research agenda," *International Journal of Information Management*, vol. 53, no. 8, p. 102104, 2020.
13. S. Robinson, C. Orsingher, L. Alkire, A. D. Keyser, M. Giebelhausen, K. N. Papamichail, P. Shams, and M. S. Temerak, "Frontline encounters of the AI kind: An evolved service encounter framework," *Journal of Business Research*, vol. 116, no. 8, pp. 366–376, 2020.
14. B. C. Stahl, "Responsible innovation ecosystems: Ethical implications of the application of the ecosystem concept to artificial intelligence," *International Journal of Information Management*, vol. 62, no. 2, pp. 1–13, 2022.
15. R. Toorajipour, P. Oghazi, and M. Palmié, "Data ecosystem business models: Value propositions and value capture with artificial intelligence of things," *International Journal of Information Management*, vol. 78, no. 10, pp. 1–23, 2024.
16. M. Weber, A. Hein, J. Weking, and H. Kremer, "Orchestration logics for artificial intelligence platforms: From raw data to industry-specific applications," *Information Systems Journal*, vol. 35, no. 3, pp. 1015–1043, 2025.
17. J. Wirtz, W. H. Kunz, N. Hartley, and J. Tarbit, "Corporate digital responsibility in service firms and their ecosystems," *Journal of Service Research*, vol. 26, no. 2, pp. 173–190, 2023.

Publisher's Note: The publisher remains impartial concerning jurisdictional claims in published maps and institutional affiliations. Responsibility for the content rests entirely with the authors and does not necessarily reflect the publisher's perspectives.