International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Privacy Preservation and Personal Data Deletion: State-of-the-Art Review and a Proposed LLM-Based Multi-Agent Architecture

Heeba Deshmukh[1], Hemali Dholam[2], Sneha Ingale[3], Dr. Vidya Sagvekar[4]

[1,2,3]Student, Artificial Intelligence and Data Science, K.J Somaiya Institute of Technology
[4]Professor, Artificial Intelligence and Data Science, K.J Somaiya Institute of Technology

Abstract
Privacy policies are the primary legal documents through which online platforms disclose their data practices, yet they remain difficult for users to understand and even harder to act upon. Parallel research has explored automated privacy policy analysis using natural language processing (NLP), large-scale web privacy measurement through browser instrumentation, and the emergence of agentic large language model (LLM) systems capable of tool use and decision making. However, these research directions largely exist in isolation. This review paper synthesizes advances across privacy policy Natural Language Processing (NLP), web privacy measurement frameworks, Retrieval - Augmented Generation (RAG) and knowledge-graph-enhanced retrieval for long and complex documents, agentic LLM architectures that combine reasoning with tool execution, and guardrails for safe autonomous agents. However, across these separate research threads, we identify a lack of any actionable direction for users to be able to control their own data. The paper highlights how the convergence of these technologies enables the design of an agentic privacy assistant that can interpret privacy policies, verify platform behavior against those policies, and autonomously execute user-authorized privacy actions such as data deletion; and outlines the architectural principles derived from current research to realize such a system.

Keywords: Agentic AI, Privacy Policy Analysis, Retrieval-Augmented Generation (RAG), Knowledge Graph Reasoning, Web Privacy Measurement, OpenWPM, LLM Guardrails, Tool-Using Language Models, Compliance Verification, Automated Data Deletion, Multi-Agent Systems, Privacy-Preserving AI.

1. Introduction

Privacy policies are official disclaimers used by websites to inform users how their information will be collected, used, shared, and stored. While legally important, they are verbose, legalistic, and require expert knowledge to comprehend.

Research has investigated the use of natural language processing to extract information from and query privacy policies, and their automatic classification [2], [4], [5]. LLMs' ability to accurately extract privacy practices without large training datasets [7] and serve as policy interpreters [17] has also been

demonstrated. And, Retrieval-Augmented Generation (RAG) has shown to reduce LLM hallucinations and improve their factual accuracy by grounding their output on external documents [1], [12].

Studies on web privacy measurement revealed that observable tracking behaviors, like cookies, fingerprinting and third-party calls, can be reliably measured using frameworks such as OpenWPM [19]. Further, a substantial body of work has been devoted to the development of agents that can understand, reason over, and act upon external resources through various tool-use frameworks [8], [9], [10], [13].

Despite these significant advances in individual fields, the knowledge and means to control their data and enforce their right to privacy remain inaccessible for the general public. This review surveys relevant literature to propose a system that attempts to integrate all the capabilities outlined above into a unified system capable of understanding privacy policies, verifying adherence through actual behavior, and automating execution of user authorized actions for their privacy.

2. Literature Review

A. NLP in Understanding Privacy Policies

The creation of fundamental datasets such as OPP-115 [5] has contributed to the ability to classify privacy practices. From this point, the domain has advanced to question-answering approaches which link user queries to relevant policy statements [2], [4]. In Polisis, scalable automated analysis was performed through hierarchical classifiers and QA pipelines [6]. With LLMs, even without fine-tuning, the prompt-based extraction of privacy practices can be performed with very high accuracy [7], and studies reveal LLMs to be better than humans at reading long privacy policies with the correct prompt [17]. All of these studies exhibit the viability of performing policy understanding at scale.

B. Behavior Measurement in Privacy

The development of OpenWPM enables the ability to actually collect evidence of real website tracking at scale [19]. The ability to gather this evidence is the basis for comparison with policy claims and is the key to auditing compliance.

C. Long Documents, RAG, and GraphRAG

Works like RAG [1] and PaperQA [12] were able to achieve reliable long document understanding through grounding with external evidence. In order to manage long policy documents which often reference terms in many other locations, GraphRAG was introduced which enables multi-hop reasoning on the document using knowledge graphs [20].

D. From Interpretation to Compliance Auditing

PrivComp-KG is a knowledge-graph and LLM-based auditing system that compares stated policy clauses to collected tracking data [18]. However, this tool primarily performs detection and does not include any form of actionable responses.

E. Tool-Using LLMs and LLMs that browse the web
The emergence of ReAct [8], Toolformer [9], WebGPT [10], Tool Learning [11], Gorilla [14], and AutoGPT [13] clearly shows that LLMs are not simply text generators but that they are able to perform multiple complex tasks by interacting with the environment.

F. Agent Architectures and Multi-Agent Systems
In order to discuss the requirements and capabilities of agent-based LLMs, an overview [15] and architecture [23] are needed to generalize and explain their design. AgentNet [20], Agent-as-a-Graph [24], and collaborative multi-agent RAG [25] show ways for many specialized agents to work together to solve tasks with graph orchestration.

G. Guardrails and Privacy Preserving Systems
Guardrail frameworks help to validate LLM actions and outputs before tool calls. The importance of data leakage prevention when dealing with user credentials and legal processes in privacy systems necessitates secure RAG and multi-agent architectures [16], [22].

3. Research Gap

The reviewed literature has demonstrated that considerable effort has been placed in related fields in automated privacy. Specifically, for the application of NLP for privacy policy, it has been demonstrated that legal text on privacy can be understood by machines. This has been achieved in multiple fashions such as through QA models [7], [17], annotated corpora [2]-[6] and more recently prompt-based reasoning on LLMs [7], [17]. Meanwhile, measurement systems such as OpenWPM have become excellent at measuring Web privacy at scale in the real world [19]. Using RAG and GraphRAG, an ability to reason over long, complex documents such as privacy policies has been created [1], [12], [20]. Similarly, recent advances have been made in agentic LLMs: language models that can reason, use tools, browse the web, call APIs and complete tasks on their own [8]-[15], [23]-[25]. For safe, correct, and non-leaking use cases of agentic LLMs, guardrails have been implemented [16], [18], [22].

However, the current research works are largely unconnected:

- Use Privacy policy NLP studies privacy policy understanding, but not acting on them [2]-[7].
- Web privacy measurement addresses the need to observe behavior of the website in practice, but does not verify or compare against stated policies [19].
- Compliance research, such as PrivComp-KG, mainly concerns discrepancies, not user-authorized remediation [18].
- Agentic LLM research focuses on tool usage and autonomy, but lacks domain-specific features [8]-[15].

Currently, no single work has explored a way to allow the general users of these websites to understand and utilize these technologies, to exercise their right of privacy and control of their data.

4. Findings from Literature Survey

The literature review reveals several consistent findings:

- QA-based systems (LLM-based systems included) are capable of reliably extracting structured data from privacy policies [2], [4], [7], [17].
- OpenWPM offers a scalable and replicable framework that collects large-scale data on actual tracking behavior [19].
- RAG is known to reduce LLM hallucination and enhance document groundedness [1], [12].
- GraphRAG enables reasoning across multiple scattered document sections for long documents [20].
- LLMs are capable of using tools correctly, performing searches and invoking APIs using ReAct, Toolformer, WebGPT, and Gorilla [8]-[14].
- Multi-agent systems can provide scalability and domain-specificity for complex workflows [20], [24], [25].

These findings demonstrate that all technical components required for such an autonomous privacy assistant already exist separately.

5. Proposed System Design

The existing technical capabilities necessary for an agent-based privacy assistant have been discussed. These are fragmented across the literature, and are synthesized below into a cohesive agent-based architecture. The following is an architectural concept, and not an implementation claim.

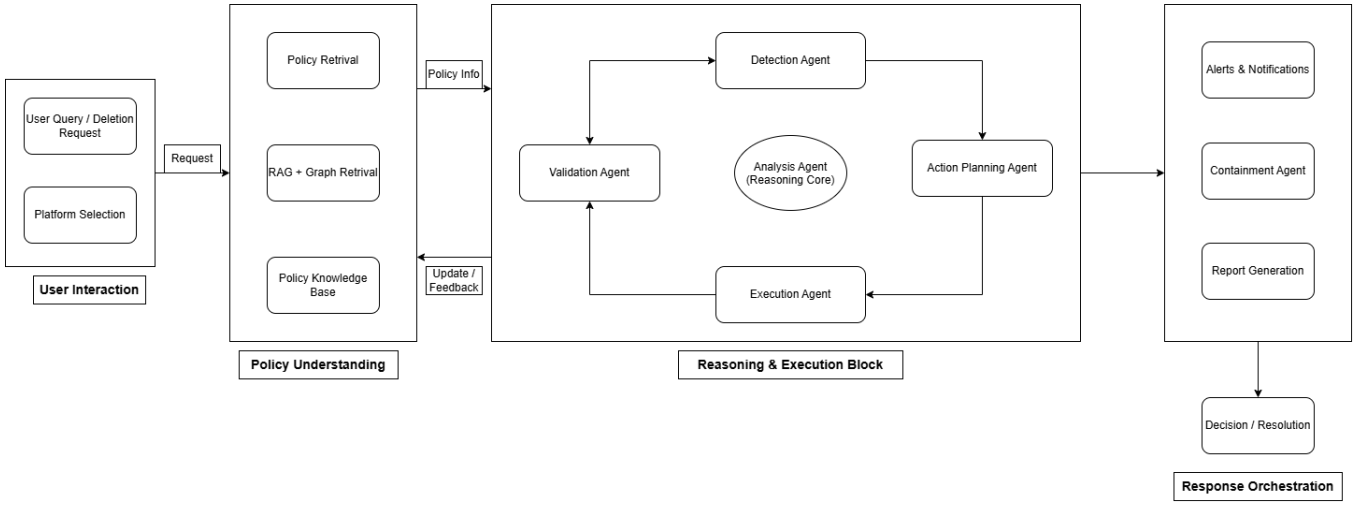


Fig. 1. Multi-Agent Workflow of the Proposed System

This diagram summarizes a layered, agent-orchestrated system that consists of four main functional blocks. These are four composite functions working together in a pipeline, and are informed by the relevant literature discussed herein.

User Interaction is the input stage of the system. User inputs, like a user's platform-specific inquiry of a data practice or a deletion request for a platform are processed at this stage.

Policy Understanding acts as a document intelligence module for the system. The Policy Agent retrieves and populates a policy knowledge base.

The Reasoning & Execution layer serves as the "agentic core" of the system. It is comprised of a Detection Agent that tracks the data collection of a platform; an Analysis Agent that compares knowledge of structured policies and observed platform behavior for policy compliance breaches; an Action Planning Agent that creates an action plan for user requests based on retrieved data; an Execution Agent that implements the planned steps; and a Validation Agent that verifies each action at the execution level.

The Response Orchestration layer is responsible for delivering information of the work done to the user, including updates on requested tasks and a full report of the system actions taken and their outcome in a human-readable format.

In summary, this architecture lays out a plan to build an end-to-end system that can interpret privacy policies, find violations of policies, and delete user data automatically using document retrieval, graph-based reasoning and agent-based execution, supported by the research in the above literature survey.

6. Conclusion

This review paper identified how several distinct lines of research, ranging from policy comprehension to web privacy measurement and the exploration of RAG, knowledge graph reasoning, agentic LLMs, and the autonomous use of tools and web browsers by agents, can be unified into a new category of system: an agentic privacy assistant.

The literature, however, is fragmented; works on policy comprehension, web privacy measurement, retrieval and reasoning over information or tool and browser interaction and trustworthy AI guardrails each operate in their own domains. No end-to-end system has yet combined these functionalities to interpret privacy policy, compare real-world system behavior against it, or execute policy-driven privacy actions like data deletion. The unification process taken in this paper thereby identified this integration gap, proposing an agent architecture from existing work.

The proposed architecture comprises multiple agents, each one working on a distinct approach to exercising user privacy.

The Policy Agent answers user queries by retrieving from and adding to a policy knowledge base. The Detection Agent tracks the data collection practices of a platform; an Analysis Agent that compares a platform’s privacy policy and their actual observed behaviour for compliance breaches; an Action Planning Agent that plans the necessary steps to be taken for a user request based on retrieved data; an Execution Agent that implements the planned steps; and a Validation Agent that verifies each action for user consent and policy basis.

This reveals that these previously disparate lines of study can be united, opening up new areas for research on verifiable legal interpretation, evidence collection at scale, safe tool operation, and agent privacy-preserving memory and evaluation.

In sum, the relevant literature now contains all necessary components to realize an intelligent system which helps users to exercise their privacy rights. Integration, rigorous testing and refinement are the tasks that the research community now needs to accomplish to turn such agents into useful, trustworthy, privacy-preserving tools.

References

1. P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, "Retrieval-augmented generation for knowledge-intensive NLP tasks," Advances in Neural Information Processing Systems (NeurIPS), vol. 33, pp. 9459–9474, 2020.
2. W. Ahmad, J. Chi, Y. Tian, and K.-W. Chang, "PolicyQA: A reading comprehension dataset for privacy policies," Findings of the Association for Computational Linguistics: EMNLP, pp. 743–749, 2020, doi: 10.18653/v1/2020.findings-emnlp.66.
3. A. Shankar, A. Waldis, C. Bless, M. A. Rodriguez, and L. Mazzola, "PrivacyGLUE: A benchmark dataset for general language understanding in privacy policies," Applied Sciences, vol. 13, no. 6, p. 3701, 2023, doi: 10.3390/app13063701.
4. A. Ravichander, A. W. Black, and N. Sadeh, "Question answering for privacy policies: Combining computational and legal perspectives," Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP-IJCNLP), pp. 4949–4959, 2019, doi: 10.18653/v1/D19-1500.
5. S. Wilson, F. Schaub, A. Dara, F. Liu, S. Cherivirala, M. Leon, et al., "The creation and analysis of a website privacy policy corpus," Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL), pp. 1330–1340, 2016, doi: 10.18653/v1/P16-1126.
6. H. Harkous, K. Fawaz, R. Shin, and K. Aberer, "Polisis: Automated analysis and presentation of privacy policies using deep learning," 27th USENIX Security Symposium (USENIX Security), pp. 531–548, 2018.
7. D. Rodriguez, I. Yang, J. M. Del Alamo, and N. Sadeh, "Large language models for privacy policy analysis: Opportunities and challenges," Computing, vol. 106, pp. 3879–3903, 2024, doi: 10.1007/s00607-024-01331-9.
8. S. Yao, J. Zhao, D. Yu, N. Du, I. Shafran, K. Narasimhan, and Y. Cao, "ReAct: Synergizing reasoning and acting in language models," arXiv preprint arXiv:2210.03629, 2022.
9. T. Schick, J. Dwivedi-Yu, R. Dessì, R. Raileanu, M. Lomeli, L. Zettlemoyer, et al., "Toolformer: Language models can teach themselves to use tools," Advances in Neural Information Processing Systems (NeurIPS), 2023.
10. R. Nakano, J. Hilton, S. Balaji, J. Wu, L. Ouyang, C. Kim, et al., "WebGPT: Browser-assisted question-answering with human feedback," arXiv preprint arXiv:2112.09332, 2021.
11. X. Qin, Y. Wang, and W. Chen, "Tool learning with foundation models," arXiv preprint arXiv:2304.08354, 2023.

12. J. Lála, O. O'Donoghue, A. Shtedritski, S. Cox, S. G. Rodriques, and A. D. White, "PaperQA: Retrieval-augmented generative agent for scientific research," arXiv preprint arXiv:2312.07559, 2023.
13. T. Richards, "AutoGPT: Autonomous GPT-4 experiments," GitHub Repository, 2023. [Online]. Available: https://github.com/Significant-Gravitas/Auto-GPT
14. S. Patil, T. Zhang, X. Wang, and J. E. Gonzalez, "Gorilla: Large language model connected with APIs," arXiv preprint arXiv:2305.15334, 2023.
15. X. Wang, Y. Li, and Z. Zhang, "Large language model-based agents: A survey," arXiv preprint arXiv:2308.11432, 2023.
16. C. Agarwal, N. Erukulla, and R. Gupta, "Privacy-Preserving LLM Infrastructure With Multi-Agent Orchestration and RAG-Driven Retrieval," Journal of International Crisis and Risk Communication Research, vol. 9, no. 1, pp. 133–140, Jan. 2026, doi: 0.63278/jicrcr.vi.3596.
17. P. Pałka, F. Lagioia, R. Liepina, M. Lippi, and G. Sartor, "Make privacy policies longer and appoint LLM readers," Artificial Intelligence and Law, 2025, doi: 10.1007/s10506-025-09442-0.
18. L. Garza, L. Elluri, A. Kotal, A. Piplai, D. Gupta and A. Joshi, "PrivComp-KG: Leveraging Knowledge Graph and Large Language Models for Privacy Policy Compliance Verification," arXiv preprint arXiv:2404.19744, Apr. 2024.
19. S. Englehardt, C. Eubank, P. Zimmerman, D. Reisman and A. Narayanan, "OpenWPM: An Automated Platform for Web Privacy Measurement," Proc. Workshop on Technology and Consumer Protection, Mar. 2015.
20. Y. Yang, H. Chai, S. Shao, Y. Song, S. Qi, R. Rui, and W. Zhang, "AgentNet: Decentralized Evolutionary Coordination for LLM-based Multi-Agent Systems," arXiv preprint arXiv:2504.00587, Apr. 2025.
21. H. Han, Y. Wang, H. Shomer, K. Guo, J. Ding, Y. Lei, M. Halappanavar, R. Rossi, S. Mukherjee, X. Tang, Q. He, Z. Hua, B. Long, T. Zhao, N. Shah, A. Javari, Y. Xia and J. Tang. "Retrieval-Augmented Generation with Graphs (GraphRAG)", arXiv:2501.00309v2, Jan. 2025.
22. M. Panda and S. Mukherjee, "Advancing Privacy and Security in Generative AI-Driven RAG Architectures: A Next-Generation Framework," International Journal of Artificial Intelligence & Applications, vol. 16, no. 2, Mar. 2025, doi: 10.5121/ijaia.2025.16203.
23. M. A. Ferrag, N. Tihanyi, and M. Debbah, "From LLM Reasoning to Autonomous AI Agents: A Comprehensive Review," arXiv preprint arXiv:2504.19678, Apr. 2025.
24. F. Nizar, E. Lumer, A. Gulati, P. H. Basavaraju, and V. K. Subbiah, "Agent-as-a-Graph: Knowledge Graph-Based Tool and Agent Retrieval for LLM Multi-Agent Systems," arXiv preprint arXiv:2511.18194, Nov. 2025.
25. A. Salve, S. Attar, M. Deshmukh, S. Shivpuje, and A. M. Utsab, "A Collaborative Multi-Agent Approach to Retrieval-Augmented Generation Across Diverse Data," arXiv preprint arXiv:2412.05838, Dec. 2024.