ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 02 Issue: 02
RECEIVED: APRIL 23 REVISED: MAY 05 ACCEPTED: MAY 27 PUBLISHED: JUNE 21

Governed AI Agents for Secure Enterprise Automation

Pierre Moreau
Asst Professor

Abstract

Governed Enterprise Intelligence – the monitored use of information by organizations for operational advantage – is crucial to navigating current turbulence affecting enterprises. Large language model (LLM)-driven agent-based automation is an enterprise-intelligent approach facilitating optimal information use. However, due to LLM vulnerabilities, spontaneous agent automation poses security and governance challenges that risk undermining their operational advantages. Addressing these challenges greatly increases agent operational performance while deploying secure services.

An LLM-driven agent is an autonomous software application capable of engaging within human-language dialogue and supporting humans across inquiry and action tasks. Agents are outfitted with mission statements and executed plans drawn from coherent information, which enable reasoning and action across simple and complex sets of sub-tasks. Complex tasks are achieved through multi-Agent collaboration, and Agents source and verify external service information from human users and available systems.

Keywords : Governed Enterprise Intelligence, Generative Agents, Policy-Driven Automation, Information Technology Governance, Security-by-Design, Responsible AI, Risk Assessment, Secure-by-Design Solutions, Technology Risk Management.

1. Introduction

Organizational assets must be actively monitored and protected against legal violations, security breaches, and other risks that could lead to financial damage or reputation loss. Consequently, a growing number of firms have started establishing Managed Security Service Providers (MSSPs) and investing in Managed Detection and Response (MDR) capabilities. Although automation improves efficiency, yet, many organizations still lack automation for critical security operations. Security-oriented automation is therefore essential, not only to enhance efficiency but also to align operational outcomes with corporate standards and decision makers' choices. However, existing automation solutions tend to focus on generic problems and solutions instead of tailoring decision processes to fit corporate standards and business operations.

Enterprise Intelligence (EI) refers to the capability of an enterprise to deploy the right assets in the right place at the right moment. In other words, it is the degree of alignment between the actual asset utilization and the situation in which they would be the most beneficial. Due to active economic pressure on the cost-effectiveness of enterprise operations, policies that improve operational efficiency and effectiveness—e.g., encouraging asset usage among decision makers and enterprise managers—are gaining crucial importance. Since a large number of records remain underutilized, designing different services that bundle those records into well-defined assets may soon become the best alternative to the costly maintenance of a large number of underutilized assets.

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 02 Issue: 02
RECEIVED: APRIL 23 REVISED: MAY 05 ACCEPTED: MAY 27 PUBLISHED: JUNE 21

1.1. Background and Significance

Governed Enterprise Intelligence is emerging as a necessity for all organizations, driven by both external and internal pressures. It is understood as the ability to detect, understand, predict, and exploit sources of change in the environment for competitive advantage while supporting compliance with policies, regulations, and laws. Governments, regulatory agencies, industry groups and the market are all pressing organizations to comply with the laws they set or to at least move in the direction of complying with them; meeting customer and market expectations for safety, quality, privacy, and sustainability; and considering the impact of their actions on society, the economy, the environment, and future generations. Institutions are increasingly judged not only by their actions but also by the actions of their direct and indirect business partners and supply chain partners.

Enterprise Intelligence covers disparate aspects of Change Detection, Knowledge Management, and Anticipation Analysis. Some parts of Organizations are automating their response to routine service requests via Large Language Models (LLMs) and LLM-driven Assistant and Conversational Agent technologies. Automated Responses and Service Fulfilments, however, remain a concern for organizations that have embraced Enterprise Intelligence or an equivalent concept, given the critical importance of Governed AI. During the automation of Request to Fulfilment Service Pipelines, it should be ensured that the automated responses are consistent with the policies and regulations of the enterprise.

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 02 Issue: 02
RECEIVED: APRIL 23 REVISED: MAY 05 ACCEPTED: MAY 27 PUBLISHED: JUNE 21

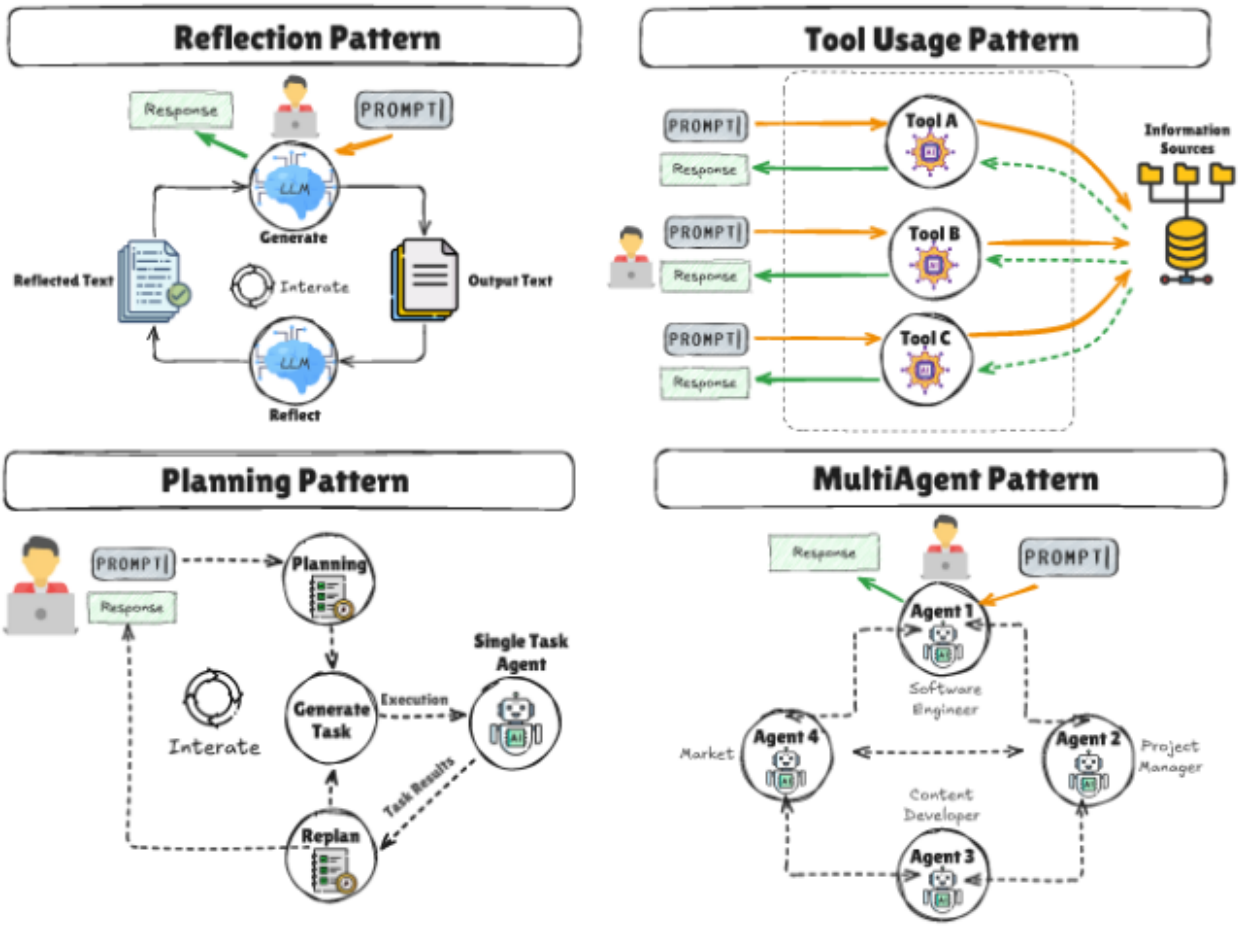


Fig 1: LLM-Driven Agents and RAG Architectures

1.2. Research design

Historically grounded in the development and deployment of digital solutions targeting improved user-centric and organizational decision-making, the emergence of large language models is poised to reduce the technical and cognitive barriers to intelligent service automation based on human-computer interaction. However, embracing these developments without a corresponding focus on security, control, and accountability would be an error of judgment for any responsible organization. Security-by-design supports the automation of LLM-based agents with the appropriate safeguards, mitigating exposure to data leakage, availability, and integrity threats. Risk-assessed governing frameworks covering policy-driving automation enable organizations to establish, communicate, and maintain responsibility over incident and process decisions.

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 02 Issue: 02
RECEIVED: APRIL 23 REVISED: MAY 05 ACCEPTED: MAY 27 PUBLISHED: JUNE 21

The goal is to explore the implications of fundamentally challenging transformation of LLM technology—the application of agents enriched by LLMs as operational interfaces for end-to-end service automation—especially from the often-overlooked angles of governed enterprise readiness and security. The study maps the emerging agent automation landscape to identify focus areas, formulates security-by-design principles to support threated assessments and validates how applied to an organization succeeds a policy-compliant deployment. The research design section has two parts: the first outlines the study's overarching aim of enabling the secure industrial exploitation of LLM-driven agents, while the second provides a shorter account of the research landscape, gaps, and contribution.

Equation 1: Asset alignment score

For asset i, mismatch $e_i = |x_i - b_i| / b_i$
Alignment for asset i: $a_i = 1 - e_i$
Weighted enterprise alignment: $A = (\Sigma w_i a_i) / (\Sigma w_i)$

Step: The article defines enterprise intelligence as deploying the right assets in the right place at the right moment. That means intelligence rises when actual use x_i approaches the beneficial use b_i.

Step: A normalized mismatch e_i is created so all assets can be compared on the same scale regardless of size.

Step: Because alignment should improve as mismatch decreases, define $a_i = 1 - e_i$.

Step: Assets do not matter equally, so combine them through a weighted average.

Interpretation: A close match between actual and beneficial utilization makes A approach 1; poor allocation drives A toward 0.

2. Background and Motivations

Adapting concepts from software engineering, Enterprise Intelligence (EI) exploits system technologies to execute business processes or tasks. The idea is thus similar to that of Business Process Automation. However, EI goes further, enabling the automation of higher-order business and strategic tasks and supporting both operational and tactical services. It considers both the systems used for executing these processes and the organisation's context in which it operates: business goals, mission, values, policies, available technologies, and so forth. Indeed, organisational governance is critical for a successful EI deployment. Governance shapes the boundaries of what can or cannot be done, the resources that can or cannot be used, the way decisions are made, and defines the delegation of responsibilities. Therefore, to properly guide adoption and implementation, EI must encompass not just the availability of advanced technological instruments but also appropriate governance.

Such governance is essential. Governance entails setting the rules for "who can do what, when, and in what conditions" in a controlled and consistent way. Ensuring that actions are compliant with corporate policy is essential for an enterprise. Recent years have seen the emergence of LLM (Large Language Models)-driven intelligent agents. These systems, empowered by the latest advances of AI technologies, can autonomously perform a wide variety of tasks: planning, reasoning, and acting. They are great enablers for enterprise service automation. However, their abilities are not self-contained. To be used effectively within the boundaries defined by the organisation and in a policy-compliant way, they need to be supervised by the concerned actors for each execution. This presents a great opportunity for the governance of intelligent agents: ensuring "authentication, authorisation, and auditing" when performing actions and discovering information.



Equation 2: Security-by-design gate for any action

Let a, u, p, c ∈ [0,1] denote authentication, authorization, provenance, and confidentiality sufficiency.
Define action admissibility score: $D = a \times u \times p \times c$
Decision rule: execute only if $D \geq \tau$, where τ is the minimum governance threshold.

Step: The article repeatedly centers authentication, authorization, provenance, and confidentiality as mandatory safeguards.

Step: A multiplicative model is appropriate because one missing control should sharply reduce admissibility.

Step: If any control is zero, then $D = 0$ and the action is blocked automatically.

Step: A threshold τ allows the enterprise to encode how strict the policy engine must be.

Interpretation: This equation formalizes least privilege and secure-by-design service execution.

2.1. Enterprise Intelligence and Governance

Enterprise Intelligence encompasses the knowledge and decision-making aptitude of an organization. Together with associated systems, these concepts have been part of information systems (IS) research for over two decades. Enterprises are governed by their internal and external stakeholders and in particular, decisions on risk, compliance, auditing, and data privacy, among others, enable intelligent automation without compromising an organization's goals. The increasing popularity of agent technologies, and in particular agent architectures based on Large Language Models (LLM), have enabled improved intelligent automation. These architectures exhibit capabilities that range from service planning and delivery to more advanced reasoning approaches, including what could be qualified as workshop-like scenarios wherein agents combine and elicit information from human, software, and organizational knowledge resources to perform an action or solve a problem.

Recognizing the need for timely and responsible automation of operational support services in any organization, a company's decision stakeholders expect that such activities can and should be ruled by stringent security and governance controls, just as any development of new systems or software automation. The proposed lines of investigation address governance for intelligent automation per se, focusing on two managerial aspects: the governance of LLM-driven agents that perform automation as the main task and intelligent automation with Retrieval-Augmented Generation (RAG) architectures when policy compliance of the overall control or support service is critical.

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 02 Issue: 02
RECEIVED: APRIL 23 REVISED: MAY 05 ACCEPTED: MAY 27 PUBLISHED: JUNE 21

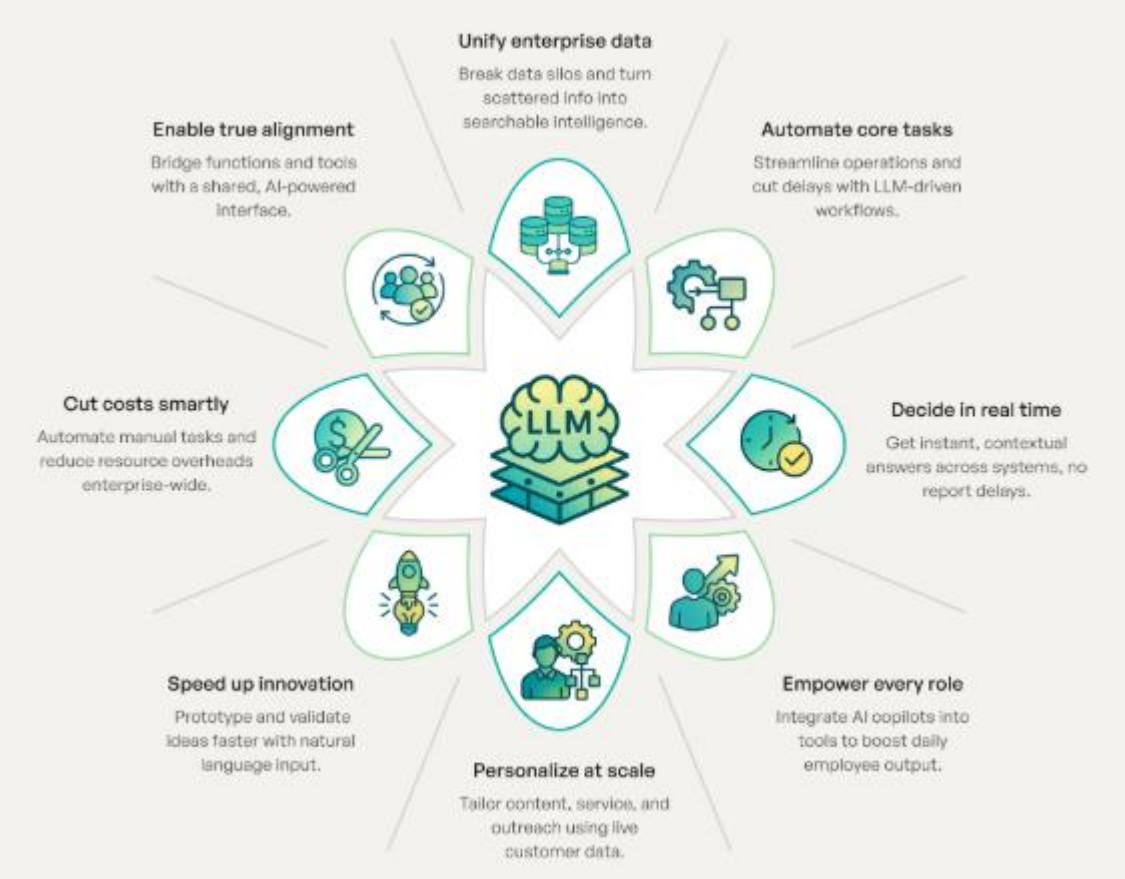


Fig 2: Business needs an enterprise LLM

2.2. Research Summary

Governed enterprise intelligence infers operational understanding from heterogeneous data, employing trusted, autonomous, cyber-physical agents for service delivery. These agents, driven by large language models to enhance planning and reasoning, direct requests within a secure ReAct-Choose-Act pipeline, permitting execution by specialized agents. Ensuring security, compliance, and operational readiness demands appropriate design during both agent engineering and enterprise governance assessment.

3. LLM-Driven Agents for Service Automation

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 02 Issue: 02
RECEIVED: APRIL 23 REVISED: MAY 05 ACCEPTED: MAY 27 PUBLISHED: JUNE 21

Large Language Model (LLM)-driven agents are increasingly invoked to address common enterprise needs with little or no human involvement. Agents are cognitive systems capable of acting within dynamic and information-rich environments. An agent-based approach can eliminate the need for humans in the execution of simple and repetitive tasks, extending the reach of large language models to all enterprise users and uses while remaining secure.

Agent architectures can come in a variety of forms, consisting of any combination of modules capable of planning, reasoning, acting, and learning on behalf of a user. Common roles found within agent systems include a planning engine for autonomously generating courses of action in human-readable language, a service orchestrator enabling the sequential execution of tasks across multiple different services, and an external executor for the translation of natural language requests into a form understandable by existing service APIs. When considered together, agents enable the end-to-end fulfillment of tasks spanning multiple enterprises and enterprise boundaries, irrespective of whether the agent itself is able to drive execution of the task's individual components.

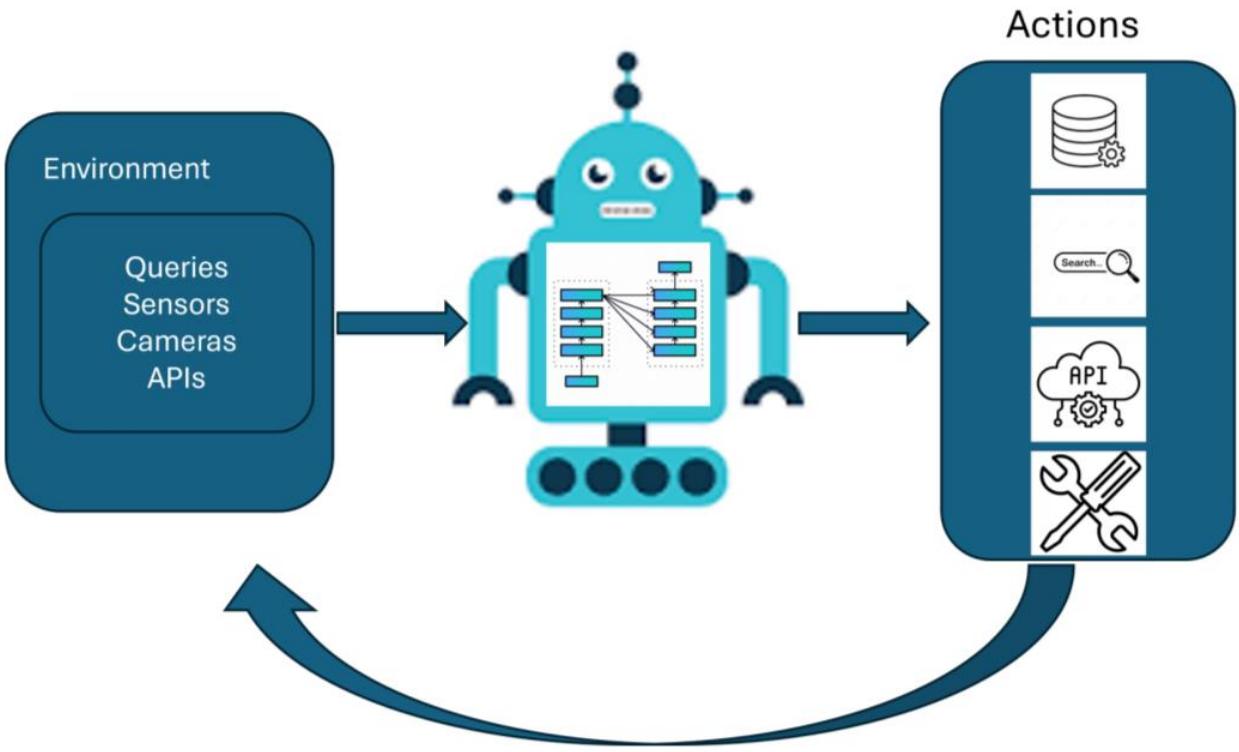


Fig 3: LLM-Powered AI Agent Systems

3.1. Agent Architectures and Capabilities

Two distinct components comprise the interaction pattern of an agent architecture situated within a service automation domain. A user or an agent itself issues a command to the agent, which subsequently proposes a plan for executing it. Planning, reasoning

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 02 Issue: 02
RECEIVED: APRIL 23 REVISED: MAY 05 ACCEPTED: MAY 27 PUBLISHED: JUNE 21

about the expected outcome, and action execution unfold in the flow of the agent interaction dominance structure, thereby assigning the tasking function of the agent to an external entity. The planning function verifies the correctness of the task definition and issues a command for the agent to act. The sequence supports a multi-agent setting where volunteering agents can take on a task defined by another agent. The proposed plan suffices for single-agent operations, where planning is subsumed by the command issue. Planning enables complex behaviours.

Action execution integrates an actor capable of intervening in the execution system and managing both cooperative actions and failures. The capability expresses various other qualities, including action recognition, knowledge instantiation, model learning, model understanding, insurance simulation, and physical system authority, at a finer level of granularity. It is indeed understood as the ability to prepare and validate the means of taking an action. A user or another agent assigns a behaviour through a request that entails performing the task and evaluates that the result matches the expectation in order to regenerate or modify its model. An actor might simulate a critical step in a behaviour and gauge hazard, risk, or performance. The means to support a defined set of behaviours—at any level of hierarchy—is central, since it determines the set of scenarios that an agent can independently manage.

Equation 3: Residual risk under controls

For threat j, uncontrolled risk $r_j = l_j \times i_j$
If mitigation strength is $m_j \in [0,1]$, residual threat risk becomes $r^*_j = l_j \times i_j \times (1 - m_j)$
Total residual risk: $R = \Sigma r^*_j = \Sigma l_j i_j (1 - m_j)$

Step: Risk registers typically combine likelihood and impact, so the base risk term is the product $l_j i_j$.

Step: Controls reduce but usually do not eliminate risk, which motivates the factor $(1 - m_j)$.

Step: Summing over all identified threats yields the enterprise residual risk used by governance teams.

Interpretation: This directly matches the paper's discussion of threat modeling, risk registers, and mitigation planning.

3.2. Objective of the Study

Governance patterns and security measures are critical to LLM-driven agents. Despite their efficiency, deployment without defined procedures raises security concerns: malicious actions could compromise data integrity, confidentiality, or availability. Therefore, automation requires security-by-design frameworks. First, security principles for LLM-based automation are established, addressing authentication, authorization, and auditing. Subsequently, a security assessment identifies potential threats and malicious use cases, complemented by an attack model. These results are linked to compliance with principle-based regulations such as the GDPR, PCI DSS, or ISO 27001, enabling a legally consistent agent deployment rationale while aligning it with actors' intentions at policy level.

4. RAG Architectures in Enterprise Contexts

RAG architectures can be applied in any context, including enterprises. Governance covers data custodianship, access control, provenance, and confidentiality. To mitigate the risk of an Operator Copilot LLM blindly trusting the LLM, its output can be confirmed by another generative "corrector" model. Access to data can be enforced via a RAG-style architecture. Metalanguages can validate input comprehension and completeness, and allow human supervision or control. In agent systems, careful analysis

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 02 Issue: 02
RECEIVED: APRIL 23 REVISED: MAY 05 ACCEPTED: MAY 27 PUBLISHED: JUNE 21

reveals component roles and capabilities. Planning, reasoning, and action execution are common. Governance and risk considerations for these various uses enhance security.

Restricting data access is not always possible, e.g., when planning for events not restricted externally. Data detected as malicious can be handled dynamically (e.g., volatile data locations such as Discord chat groups can see access detection). External agents influence only superficial data; core processes can still be safely performed in-house with quality assurance as appropriate. Security principles, components, configurations, and evaluations are provided for experimental service automation using these design angels. Solutions should be validated (or shown to be secure) in each actual deployment.

Equation 4: Governed enterprise intelligence objective

$$G = \alpha A + \beta S + \gamma P - \delta R$$

where $\alpha + \beta + \gamma + \delta = 1$ and $\alpha,\beta,\gamma,\delta \geq 0$

Step: The article seeks operational advantage, security, and policy compliance while reducing risk.

Step: A single scalar objective is formed by adding desirable quantities (alignment A, security S, policy compliance P) and subtracting undesirable quantity R.

Step: The coefficients α, β, γ, δ encode strategic priorities of the organization.

Interpretation: G becomes a decision score for comparing architectures, workflows, or deployment options.

4.1. Data Custodianship and Access Control

In a RAG architecture, systematic data custodianship across different data stores is a prerequisite for secure service automation using LLM-based agents. Custodianship encompasses the ability to determine the data access policies for a specific service, to authenticate the role of the agent requesting access to the data in an agent-to-store interaction, and to manage the data shareability and confidentiality conditions in an agent-to-agent interaction. When interactions take place through agent-plans, custodianship involves a higher level of intricacy.

Data custodianship is structured around the concepts of:

1. **Authentication**: The capability of a data store to verify the identity of an agent for an agent-to-store interaction, and a data store to confirm the identity of another data store for an agent-to-agent data share interaction.

2. **Authorization**: The data access policy defined in the data store for a destination agent, in the case of an agent-to-store interaction, or for an originating agent or role, in the case of an agent-to-agent interaction.

3. **Provenance**: The ability of a data store to enforce the security conditions for data sharing in an agent-to-agent interaction, and to cover the shareability condition in an agent-to-store interaction.

4. **Confidentiality**: The property of a piece of data that restricts the information about the data to a subset of the user policy.

4.2. Methodology

Reinforcement and Risk Assessment, Validation, and Threat Modelling: Evaluation of validation criteria applicable to the

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 02 Issue: 02
RECEIVED: APRIL 23 REVISED: MAY 05 ACCEPTED: MAY 27 PUBLISHED: JUNE 21

identified property classes; analysis of potential threats to the intended use of the security controls; and assessment of the control plan for deficiencies. The following outlines the discussion relevant to the set of criteria assigned to the initial class of the Cyber-Secured Generative Agent Centre of Excellence (CSGA-COE) evaluation framework. The external interactions of the RAG pipeline must underpin the security analysis. Security controls must be in place for proper authentication and authorization of the two external components—OpenID Connect operator and risk indicator LLM—responsible for controlling and enforcing the connection to the authentication management provider, thus ensuring a secure least-privilege principle, as well as for logging audit-trail information, also related to the least-privilege principle in connection with the Azure Communication Services infrastructure provider.

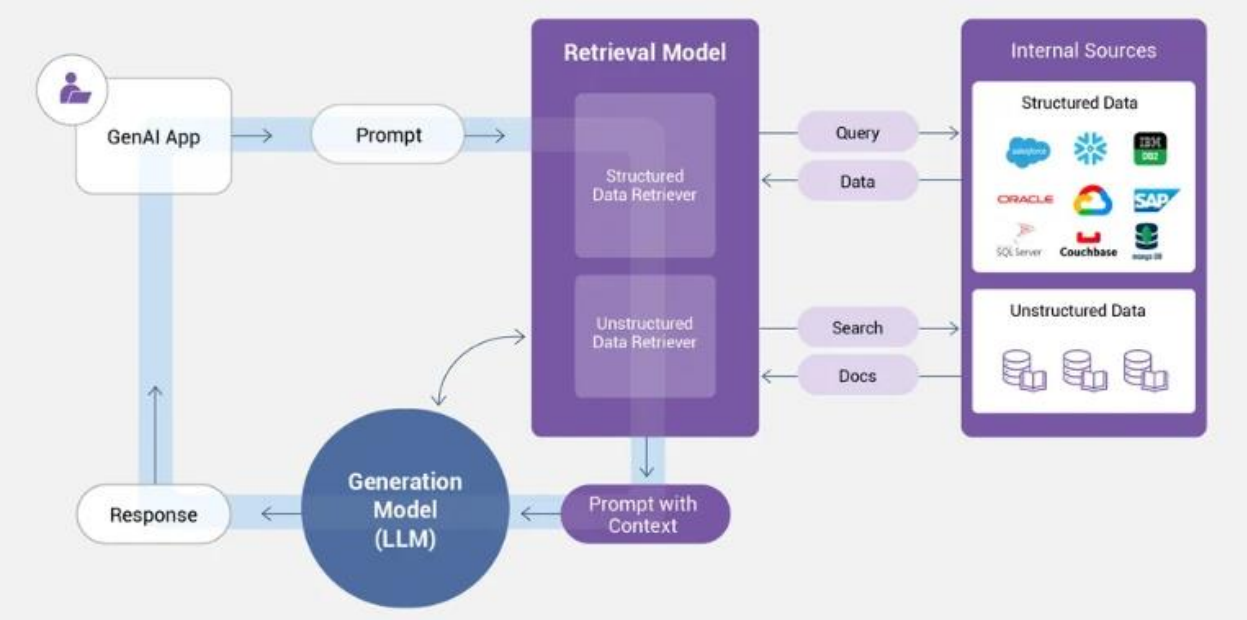


Fig 4: RAG architecture

5. Security by Design for LLM-Based Automation

For LLM-based automation, security by design entails defining authentication, authorization, and auditing mechanisms. Such mechanisms should implement least-privilege access wherever possible and ensure the necessary means of traceability for all actions taken by automated systems and LLM-driven agents. For organizations that adopt these ideas for LLM-based automation of business processes, other security-enabled architectures, concepts, and processes assist in validating the security of specific LLM-based solutions. Security assessments, threat modeling, and mapped compliance with recognized standards, e.g., ISO/IEC 27001, ISO/IEC 27018, ISO/IEC 27701, go beyond these considerations and create the necessary assurance in an organization's approach to securing LLM-based automation solutions.

Operational security must be considered when adopting LLM-based automation. Definition of the security and environment of the systems attracting data is crucial. Organizations adhering to the NIST Cybersecurity Framework for Improving Critical

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 02 Issue: 02
RECEIVED: APRIL 23 REVISED: MAY 05 ACCEPTED: MAY 27 PUBLISHED: JUNE 21

Infrastructure Cybersecurity provide a good example. Proper definition of the five functions and corresponding categories and subcategories, combined with the security urgent guidelines defined in the Executive Orders Supporting Secure Use of Artificial Intelligence, particularly Number 14091 on Further Advancing Racial Equity and Support for Underserved Communities Through the Federal Government, provides a comprehensive approach to security."

Equation 5: RAG trustworthiness score

Let q = retrieval relevance, p_f = policy fit, c_f = confidentiality preservation, a_t = auditability.
$T = w_q\, q + w_p\, p_f + w_c\, c_f + w_a\, a_t$
with $w_q + w_p + w_c + w_a = 1$.

Step: The article states that enterprise RAG must govern data custodianship, access control, provenance, and confidentiality.

Step: A weighted additive score is suitable because RAG quality depends on several complementary criteria rather than a single gate alone.

Step: The weights let an organization emphasize stronger confidentiality, stronger provenance, or stronger auditability depending on its risk posture.

Interpretation: T quantifies whether retrieved knowledge is both useful and governable.

5.1. Authentication, Authorization, and Auditing

Authentication, authorization, and auditing are fundamental security principles for LLM-based automation solutions. The key idea is that any action taken by a service executing LLM-based automation relies on these three principles to decide:

- Who is the user interacting with the automation-enabled system?

- What actions is the user allowed to perform?

- Do the actions have sufficient traces for auditing to trace back who ordered what and why?

This means that the principle of least privilege is applied: a user can perform an action only when she is allowed to and that for audit purposes, one can trace back what actions were ordered. This is also related to the notion of accountability: for each decision taken there is an identified person accountable for the decision. In addition, traceability is crucial for internal or external auditing.

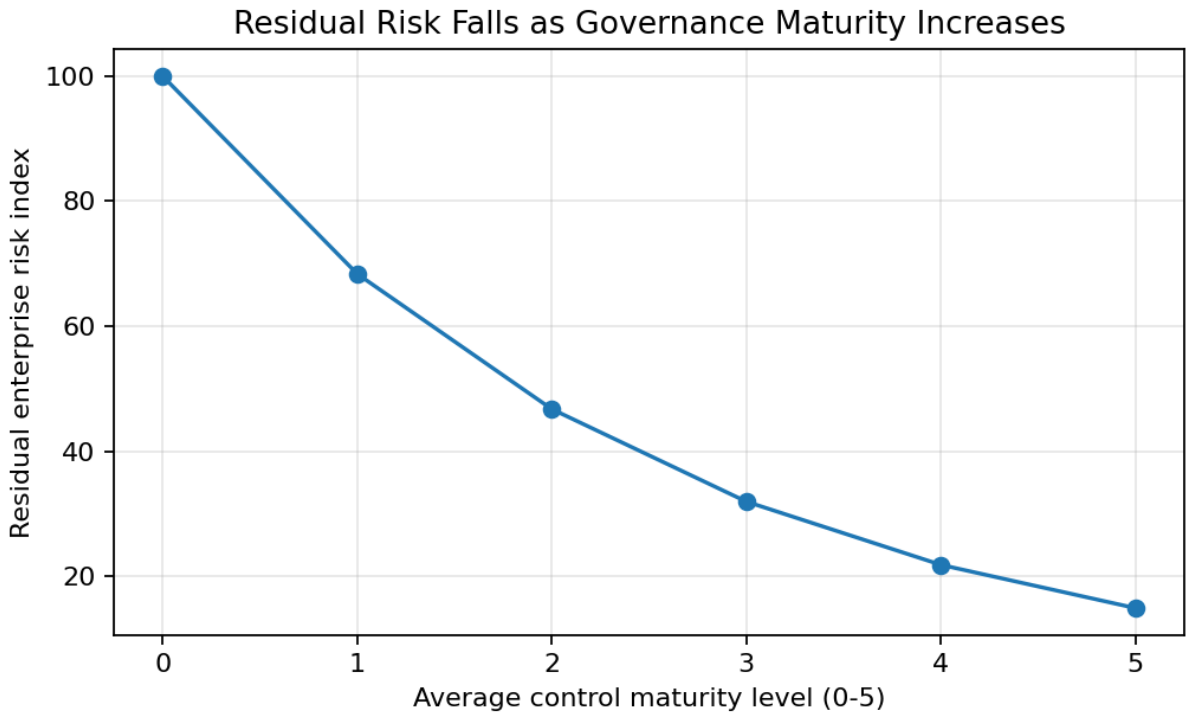
5.2. Result

Security assessments, threat modeling, testing, and maturity evaluations were carried out on three proof-of-concept agent implementations: an automated information retrieval assistant, an internal chat assistant, and a workflow automation agent with remote execution capabilities. The two assistants were found to correctly implement group-based authentication and authorization, employing the least-privilege principle and allowing detection of misuse in accordance with IT security governance policies. The workflow automation agent incorporated scheduled task operation, but lack of protection in the access to the executor controller node, coupled with excessive access privileges granted through Windows tasking, posed serious security issues that could have allowed system attack. Security threats were identified for all use cases, alongside a multi-layered

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 02 Issue: 02
RECEIVED: APRIL 23 REVISED: MAY 05 ACCEPTED: MAY 27 PUBLISHED: JUNE 21

security mitigation strategy proposal, and audit trails were incorporated to comply with auditing policies. Testing demonstrated the achievability of security-by-design automation and opened avenues for operating-managed support of LLM services.

Second, a continuous monitoring process that detects non-compliance of agent execution with the policies determined by the organization, including the auditing of agent tracking. The results of these analyses must be reported back to the organization management for corrective actions. Third, sufficient temporal separation per execution so that the risk of undetected abuse is reduced.



6. Governance Frameworks and Compliance Models

The preceding sections have separately examined the security aspects of Large Language Model (LLM)-driven agents acting on external resources within the framework of an organization and the organization-enforced information access management of Retrieval-Augmented Generation (RAG) pipelines. The completeness of these security considerations leads to a governance framework for these types of policies and an associated suggested compliance model.

The automation service execution by agents must align with the policies determined by the organization overseeing the agent execution environments. Deviations from these policies should be detected and justified; otherwise, appropriate penalization must

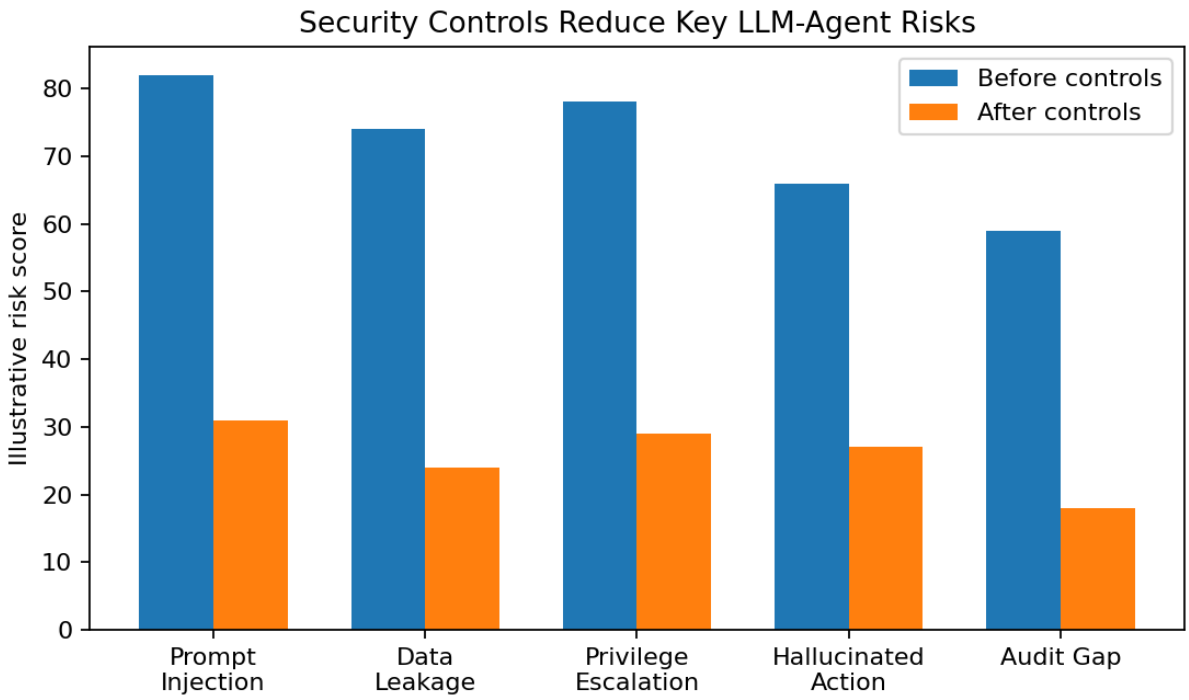
ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 02 Issue: 02
RECEIVED: APRIL 23 REVISED: MAY 05 ACCEPTED: MAY 27 PUBLISHED: JUNE 21

follow accountability processes defined by the organization. Security is consolidated by four aspects. First, independent risk assessments performed by the organization against the execution environments of the agents.

6.1. Policy-Driven Automation

Policy-driven automation aligns agent interactions with organizational policies, ensuring that agents act on behalf of the organization while fulfilling authorized tasks. An example interaction showcases a data analytics agent that acquires sensitive customer data for a targeted marketing analysis, presenting a potential infra-privacy policy violation. Such procedures require formalized policy definition and classification of decision contexts, underscoring the need to govern these agents instead of simply relying on the platform to audit their actions, which can be ineffective in preempting policy breaches.

The interaction can be classified into one of the organizational policies and sub-policies defined above. Distinct from external organizational policies that safeguard enterprise assets and reputation, internal sub-policies direct decision-makers and automata handling sensitive enterprise data. The user-assigned role to the automaton should match the agent-assigned role within the policy context. Addressing requirements related to accountability and automation easily integrates these requirements into the premises, thus driving further analysis without exploring other policy-related aspects. Learning or experience accumulation facilitates developing such models for more complex scenarios of higher strategic value for the organization.





6.2. Risk Assessment and Mitigation

Compliance with the governance framework requires risk analysis to identify risks related to automation execution and assess their potential implications. Risks are recorded in the risk register, which lists potential risks, assesses their likelihood and severity, proposes mitigating actions, designates responsible individuals, and outlines ongoing monitoring guidelines. Although the risk register initially focuses on automation utilizing LLM technologies, it should also cover other agents, detailing any additional risks and associated mitigations. Continuous maintenance of the risk register and its ongoing association with the introduced governance framework is crucial for sustaining governance throughout the development and use of automated solutions, enabling risk reassessment for individual execution situations.

Threat	Likelihood l_j	Impact i_j	Mitigation m_j	Residual r*_j
Prompt injection	0.80	0.90	0.62	0.274
Data leakage	0.70	0.90	0.64	0.227
Privilege escalation	0.60	0.95	0.61	0.222
Hallucinated action	0.65	0.75	0.59	0.200

Table : Worked numerical example

7. Conclusion

The addressed challenges and proposed solutions significantly enhance the safety and operational effectiveness of service automation in modern organizations. A pivotal aspect of enterprise intelligence is LLM-driven agent architectures, enabling the centralized management of responsibility, explainability, and risk. The potential of RAG architectures is also outlined, with a focus on governance.

8. References

1. Alomari, M. K., Khan, H. U., & Khan, S. (2021). Systematic analysis of artificial intelligence-based platforms for identifying governance and access control. Security and Communication Networks, 2021, Article 8686469.
2. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2021). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115.



3. Brundage, M., Avin, S., Wang, J., Belfield, H., Krueger, G., Hadfield, G., & Dafoe, A. (2021). Toward trustworthy AI development: Mechanisms for supporting verifiable claims. arXiv preprint arXiv:2004.07213.
4. Inala, R. (2023). Revolutionizing Customer Master Data in Insurance Technology Platforms: An AI and MDM Architecture Perspective. International Journal of Finance (IJFIN)-ABDC Journal Quality List, 36(6), 579-606.
5. Chen, C., Li, K., Duan, Z., & Chen, X. (2021). Towards responsible artificial intelligence: A review of AI governance frameworks and principles. IEEE Access, 9, 148527–148545.
6. Jobin, A., Ienca, M., & Vayena, E. (2021). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 3(5), 389–399.
7. Kaur, D., Uslu, S., Rittichier, K. J., & Durresi, A. (2022). Trustworthy artificial intelligence: A review. ACM Computing Surveys, 55(2), Article 39.
8. Aitha, A. R. (2021). Dev Ops Driven Digital Transformation: Accelerating Innovation In The Insurance Industry. Journal of International Crisis and Risk Communication Research.
9. Mökander, J., & Floridi, L. (2021). Ethics-based auditing to develop trustworthy AI. Minds and Machines, 31, 323–327.
10. Mökander, J., Morley, J., Taddeo, M., & Floridi, L. (2022). Ethics-based auditing of automated decision-making systems: Nature, scope, and limitations. Science and Engineering Ethics, 28, Article 44.
11. Mangalampalli, B. M. Generative AI Applications In Healthcare Data Mart Design And Optimization.
12. Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2021). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. Science and Engineering Ethics, 26, 2141–2168.
13. Siau, K., & Wang, W. (2021). Artificial intelligence, machine learning, automation, and robotics. Journal of Database Management, 32(1), 1–15.
14. Tabassi, E. (2021). Artificial intelligence risk management framework: Concept paper. National Institute of Standards and Technology.
15. Aravind, R., Surabhi, S. N. D., & Shah, C. V. (2023). Remote Vehicle Access: Leveraging Cloud Infrastructure for Secure and Efficient OTA Updates with Advanced AI. Eur. Econ. Lett, 13, 1309-1319.



16. Thiebes, S., Lins, S., & Sunyaev, A. (2021). Trustworthy artificial intelligence. Electronic Markets, 31, 447–464.
17. Vakkuri, V., Kemell, K.-K., Jantunen, M., & Abrahamsson, P. (2021). "This is just a prototype": How ethics are considered in software engineering practice. IEEE Software, 38(5), 63–70.
18. Inala, R. (2023). Big Data Architectures for Modernizing Customer Master Systems in Group Insurance and Retirement Planning. Educational Administration: Theory and Practice, 29(4), 5493-5505.
19. Yigitcanlar, T., Desouza, K. C., Butler, L., & Roozkhosh, F. (2022). Contributions and risks of artificial intelligence in the governance of smart cities. Smart Cities, 5(2), 900–923.
20. Zhang, D., Maslej, N., Brynjolfsson, E., Etchemendy, J., Ganguli, D., Lyons, T., Manyika, J., Niebles, J. C., Sellitto, M., Shoham, Y., Clark, J., & Perrault, R. (2022). The AI Index 2022 annual report. AI Index Steering Committee, Stanford Institute for Human-Centered Artificial Intelligence.
21. Aravind, R., Surabhi, S. N. D., & Shah, C. V. (2023). Remote Vehicle Access: Leveraging Cloud Infrastructure for Secure and Efficient OTA Updates with Advanced AI. EuropeanEconomic Letters (EEL), 13 (4), 1308–1319. Retrieved fromhttps.
22. Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., & Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.
23. Weidinger, L., Mellor, J., Rauker, T., Griffin, C., Uesato, J., Huang, P.-S., Cheng, M., Glaese, A., Balle, B., Kasirzadeh, A., Kenton, Z., Hawkins, W., Stepleton, T., Biles, C., Barnett, B., Bansal, Y., Azzam, M., Gabriel, I., Isaac, W., & Gabriel, M. (2022). Taxonomy of risks posed by language models. Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, 214–229.
24. Ganguli, D., Askell, A., Schiefer, N., Liao, T. I., Liskovich, D., & Amodei, D. (2022). Red teaming language models to reduce harms: Methods, scaling behaviors, and lessons learned. arXiv preprint arXiv:2209.07858.
25. Bai, Y., Jones, A., Ndousse, K., Askell, A., Chen, A., DasSarma, N., Drain, D., Ganguli, D., Henighan, T., Joseph, N., Kadavath, S., Kernion, J., Conerly, T., Elhage, N., El-Showk, S.,

ISSN: 3050-9696 E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 02 Issue: 02
RECEIVED: APRIL 23 REVISED: MAY 05 ACCEPTED: MAY 27 PUBLISHED: JUNE 21

El-Showk, A., Fort, S., Hume, T., Johnston, S., & Kaplan, J. (2022). Constitutional AI: Harmlessness from AI feedback. arXiv preprint arXiv:2212.08073.

26. Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., & Liang, P. (2022). On the opportunities and risks of foundation models. Communications of the ACM, 65(7), 36–47.
27. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623.
28. Weidinger, L., Uesato, J., Rauker, T., Griffin, C., Huang, P.-S., Cheng, M., Glaese, A., Balle, B., Kasirzadeh, A., Kenton, Z., Hawkins, W., Stepleton, T., Birhane, A., Gabriel, I., Isaac, W., & Amodei, D. (2021). Ethical and social risks of harm from language models. arXiv preprint arXiv:2112.04359.
29. Yandamuri, U. S. (2023). An Intelligent Analytics Framework Combining Big Data and Machine Learning for Business Forecasting. International Journal Of Finance, 36(6), 682-706.
30. Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. ACM Computing Surveys, 55(9), Article 195.
31. Mialon, G., Dessì, R., Lomeli, M., Nalmpantis, C., Pasunuru, R., Raileanu, R., Rozière, B., Schick, T., Dwivedi-Yu, J., Celikyilmaz, A., Grave, E., & Scialom, T. (2023). Augmented language models: A survey. Transactions on Machine Learning Research, 2023.
32. Privacy and security challenges for autonomous agents: A study of two social humanoid service robots. (2022). Proceedings of the 2022 IEEE International Conference on Cloud Computing Technology and Science, 272–279.