Requisitos para Concepção de Soluções de Gestão de Identidade Adequadas para IA Agêntica

Gabriel Spanhol Abreu[1], Yan Vitor Pires Pereira[1], Rodrigo Cesar Lira da Silva[2], Emerson Ribeiro de Mello[1]

[1]Instituto Federal de Santa Catarina – Câmpus São José

{gabriel.a2004, yan.pp2006}@aluno.ifsc.edu.br

mello@ifsc.edu.br

[2]Instituto Federal de Pernambuco – Câmpus Paulista

rcls@ecomp.poli.br

Resumo. *A Inteligência Artificial agêntica representa um avanço significativo nos sistemas autônomos. Entretanto, os modelos e soluções atuais de Gestão de Identidade (GId) foram concebidos, em sua maioria, para fluxos centrados em usuários humanos. Este artigo apresenta um conjunto de requisitos para orientar a concepção de soluções de GId adequadas ao contexto da IA agêntica.*

1. Introdução

Agentes de Inteligência Artificial (IA) são capazes de atuar de forma autônoma, tomar decisões e executar ações em nome de um usuário com o objetivo de cumprir tarefas delegadas [Sapkota et al. 2026]. Nesse cenário, a comunicação entre usuários e agente é bidirecional. Ao criar um agente, o usuário define tarefas e concede permissões para sua execução. Em contrapartida, o agente pode iniciar interações para solicitar dados ou ações complementares necessários ao cumprimento da tarefa atribuída [South et al. 2025a].

Em alguns casos, para que consiga executar uma tarefa, o agente personifica o usuário perante sistemas externos. Assim, tais sistemas não conseguem distinguir se a ação foi realizada pelo usuário ou por um agente que ele criou. No contexto da IA agêntica, um agente pode instanciar subagentes, compartilhando assim credenciais e permissões do usuário entre múltiplas entidades autônomas.

O principal diferencial da IA agêntica em relação a softwares e fluxos tradicionais de automação é seu comportamento flexível e não determinístico. Esses sistemas deixam de atuar apenas como *chatbots* geradores de texto e passam a operar como cargas de trabalho ativas, capazes de alterar estados locais e remotos por meio de ferramentas e da interação com ecossistemas digitais diversos [South et al. 2025b].

Modelos e protocolos atuais de gestão de identidade e de acesso foram projetados para usuários humanos e com fluxos de execução determinísticos. Nesse contexto, a personificação de usuários por agentes dificulta a rastreabilidade e a auditoria das ações executadas pelos agentes, comprometendo a responsabilização. Além disso, o uso das mesmas credenciais do usuário tende a reproduzir o mesmo conjunto de privilégios a essas associadas, contrariando o princípio do menor privilégio e tornando mais complexa a revogação da delegação. Importante ressaltar que o problema

de delegação de privilégios em computação não é algo recente, como já apresentado em [Gasser and McDermott 1990], porém, IA agêntica, devido às suas características, traz novas condições de escala, autonomia e imprevisibilidade.

Diante desse cenário, este trabalho tem como objetivos: i) apresentar características da IA agêntica diretamente relacionadas à GId; e ii) propor um conjunto de requisitos para orientar a concepção de soluções de GId adequadas a esse contexto.

2. Características da IA agêntica para soluções GId

Nesta seção são apresentadas características da IA agêntica que precisam ser consideradas por soluções de gestão de identidade e de acesso.

- **Autonomia e lacuna de determinismo**: Em arquiteturas clássicas, assume-se que cada permissão corresponde a um conjunto previsível de ações. Em agentes baseados em Grandes Modelos de Linguagem (LLM, *Large Language Models*), a execução decorre de inferência probabilística orientada por linguagem natural. Assim, a mesma permissão pode gerar comportamentos distintos conforme o contexto.
- **Efemeridade e dinamismo de ciclo de vida**: Agentes podem operar em intervalos curtos, o que torna inadequado o uso de identificadores persistentes e *tokens* de longa duração [Huang et al. 2025].
- **Delegação transversal e cadeias de confiança**: Agentes operam sob autoridade delegada, agindo em nome de usuários humanos ou de outros agentes, podendo ainda criar subagentes para delegação de tarefas [Bhushan et al. 2025]. Sistemas de GId convencionais consideram apenas uma única entidade, o usuário original. [Huang et al. 2025].
- **Identidade de agente rica e verificável**: Identidades de máquinas tradicionais, baseadas em contas de serviços e chaves opacas, são insuficientes para agentes autônomos. Em ambientes *Zero Trust*, a identidade do agente deve incluir atributos verificáveis sobre modelo, instruções, dependências e limites operacionais.
- **Interação máquina a máquina**: Agentes podem formar equipes *ad-hoc* e colaborar entre organizações sem confiança pré-estabelecida. Nesses casos, a GId deve permitir a autenticação mútua, verificação da origem e avaliação das capacidades para reduzir riscos de manipulação maliciosa e preservar a integridade das operações [Huang et al. 2025].

3. Proposta de requisitos para soluções GId adequadas para IA agêntica

A partir das características apresentadas na Seção 2 e de uma revisão narrativa da literatura, foi elaborado um conjunto de requisitos para soluções de GId adequadas para IA agêntica. Os requisitos estão organizados nas seguintes categorias e classificados como: requisitos funcionais (RF); propriedades de segurança (PS) e requisitos de governança (RG).

- **Delegação e autorização**
  - **Atenuação de escopo (PS)**: nenhum agente pode receber privilégios superiores aos do agente que o originou. Em cada etapa da delegação, as permissões devem ser preservadas ou reduzidas [South et al. 2025a, Huang et al. 2025].
  - **Delegação verificável (PS)**: A capacidade de qualquer parte interessada, seja um serviço receptor, um auditor ou o próprio sistema, poder comprovar, a qualquer momento, que uma delegação ocorreu de forma legítima, válida e devidamente autorizada, com base em evidências criptográficas ou registros verificáveis [South et al. 2025a, South et al. 2025b].

  - **Cadeia de delegação identificável (PS)**: Cada elo da cadeia deve registrar, de forma inequívoca, quem delegou, quem recebeu a delegação e sob quais condições. Assim será possível reconstruir o caminho completo da autorização, do originador ao executor final [South et al. 2025a, South et al. 2025b].

- **Escopo da delegação**
  - **Finalidade explícita da delegação (RF)**: A informação do propósito ou finalidade da delegação é importante em sistemas baseados em agentes de IA, pois estabelece de forma explícita o motivo pelo qual determinado agente foi criado e autorizado a executar uma ação específica [South et al. 2025b, South et al. 2025a, Kim and Kim 2025].
  - **Ações e recursos autorizados (RF)**: O escopo operacional do agente deve ser definido de forma explícita e completa para englobar tanto as ações que ele está autorizado a executar quanto o conjunto de recursos, APIs e ferramentas aos quais ele pode recorrer para cumprir suas tarefas. Ambas as dimensões delimitam, em conjunto, o raio de atuação permitido ao agente [Huang et al. 2025, South et al. 2025a].
  - **Granularidade temporal e contextual (RF)**: Autorizações devem incorporar restrições de tempo, contexto e condições de uso, reduzindo o risco de abuso de privilégios em operações autônomas [Huang et al. 2025, South et al. 2025b].
- **Ciclo de vida**
  - **Revogação em cadeia (RF)**: A revogação de permissões deve propagar-se por toda a cadeia de delegação, invalidando acessos derivados [South et al. 2025b, South et al. 2025a, Syros et al. 2025].
- **Auditoria e responsabilidade**
  - **Auditabilidade contínua (PS)**: O sistema deve registrar e rastrear continuamente todas as permissões concedidas, as credenciais utilizadas e as ações executadas ao longo da cadeia de delegação. Complementarmente, exige que o conjunto de permissões de cada agente seja explícito e verificável a qualquer momento, de modo que o estado de autorização do sistema seja sempre observável e auditável [Syros et al. 2025, Marchal et al. 2026].
  - **Responsabilização (PS)**: Os mecanismos existentes devem permitir a atribuição inequívoca de autoria a cada evento relevante no sistema: quem concedeu as permissões utilizadas, quem aprovou a operação e qual agente efetivamente executou a ação. A responsabilização não se restringe ao executor imediato, mas deve percorrer toda a cadeia de delegação até o autorizador original [South et al. 2025b, Marchal et al. 2026, Syros et al. 2025].
- **Governança**
  - ***Human-in-the-Loop* (RG)**: O modelo deve prever pontos de intervenção humana, opcionais ou obrigatórios conforme a criticidade, para validar operações sensíveis e reduzir riscos em decisões autônomas de grande impacto [Kim and Kim 2025, Syros et al. 2025, South et al. 2025b].
- **Interoperabilidade**
  - **Compatibilidade com sistemas existentes (RF)**: A solução deve integrar-se a infraestruturas já em uso (protocolos, serviços e mecanismos de segurança), viabilizando adoção incremental em ambientes corporativos e distribuídos [South et al. 2025b, Jones et al. 2020].

4. Conclusão

A literatura recente evidência a necessidade de mecanismos robustos de delegação, autorização e rastreabilidade em sistemas com IA agêntica. Porém, a definição dos mecanismos se encontra dispersa entre diferentes trabalhos e propostas de arquiteturas. Como principal contribuição, este trabalho organiza e estrutura um conjunto de onze requisitos essenciais como forma de orientar o desenvolvimento de soluções de gestão de identidades adequados para a IA agêntica. Embora parte desses requisitos já seja atendida por protocolos existentes, outros ainda carecem de mecanismos nativos. Como trabalho futuro, pretende-se indicar caminhos para superar essas lacunas e validar os requisitos propostos com especialistas em GId, segurança, privacidade e sistemas multiagente, quanto à sua completude, clareza e utilidade.

Agradecimentos

Este trabalho foi financiado pelo projeto Gerenciamento de Identidades e Acesso para Agentes Inteligentes, com apoio do Centro Integrado de Segurança em Sistemas Avançados (CISSA) e com recursos financeiros oriundos do PPI IoT/Manufatura 4.0 do MCTI, através do Termo de Cooperação 08/2024, firmado com a EMBRAPII.

Referências

Bhushan, B., Pappu, K., and Mittal, A. (2025). A conceptual framework for authentication in agentic AI ecosystems: Protocol analysis and taxonomy. In *2025 International Conference on Computer and Automation (ICCA)*.

Gasser, M. and McDermott, E. (1990). An architecture for practical delegation in a distributed system. In *IEEE Symposium on Research in Security and Privacy*.

Huang, K., Narajala, V. S., Yeoh, J., Ross, J., Lambe, M., Raskar, R., Harkati, Y., Huang, J., Habler, I., and Hughes, C. (2025). A novel zero-trust identity framework for agentic AI: Decentralized authentication and fine-grained access control.

Jones, M. B., Nadalin, A., Campbell, B., Bradley, J., and Mortimore, C. (2020). OAuth 2.0 Token Exchange. RFC 8693.

Kim, S. and Kim, H. (2025). A case study on delegating critical tasks to agentic ai and prototype access control methods. In *Cyber Awareness and Research Symposium*.

Marchal, N., Chan, S., Franklin, M., Revel, M., Keeling, G., Fischli, R., Chandra, B., and Gabriel, I. (2026). Architecting trust in artificial epistemic agents. arXiv:2603.02960.

Sapkota, R., Roumeliotis, K. I., and Karkee, M. (2026). Ai agents vs. agentic ai: A conceptual taxonomy, applications and challenges. *Information Fusion*, 126:103599.

South, T., Marro, S., Hardjono, T., and et al, M. (2025a). Position: AI agents need authenticated delegation. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *PMLR*.

South, T., Nagabhushanaradhya, S., Dissanayaka, A., and et al, S. C. (2025b). Identity management for agentic ai: The new frontier of authorization, authentication, and security for an ai agent world.

Syros, G., Suri, A., Nita-Rotaru, C., and Oprea, A. (2025). SAGA: A security architecture for governing AI agentic systems.