Livre blanc

La sécurité des agents d'IA, nouveau défi de l'identité

Les agents d'IA redéfinissent les expériences numériques, mais leur sécurisation implique de repenser la gestion des identités et les contrôles d'accès conçus pour l'humain.

Sommaire

Les marchés s'expriment

Le marché mondial de l'IA générative devrait passer de 67,18 milliards USD en 2024 à 967,65 milliards USD à l'horizon 2032 — un taux de croissance annuel composé de 39,6 % pendant la période envisagée.

Source : Fortune Business Insights, 2025

Le marché des agents d'IA devrait passer de 5,1 à 47,1 milliards USD entre 2024 et 2030, soit un taux de croissance annuel composé de 44,8 %.

Source : Fortune Business Insights, 2025

D'après une enquête réalisée par IBM, 59 % des entreprises travaillant déjà avec l'IA ont l'intention d'accélérer et d'accroître leurs investissements dans cette technologie, gage d'un engagement fort en faveur de l'élargissement des capacités d'IA.

Source : IBM, 2024

Les lacunes de la sécurité des identités appliquée à l'IA

L'adoption rapide des agents d'IA a engendré de nouveaux défis liés à l'identité. D'ici 2027, 82 % des entreprises devraient avoir déployé des agents d'IA. Pourtant, la plupart des stratégies de sécurité restent axées sur l'authentification humaine (source : Capgemini, 2024).
Ce décalage crée des vulnérabilités qui peuvent être exploitées par les cybercriminels. Les agents d'IA ont souvent recours à des identifiants obsolètes, ce qui en fait des cibles de choix pour le vol d'identifiants, l'usurpation d'identité et les accès non autorisés. En outre, les agents d'IA traitent un important volume de données sensibles, ce qui accroît le risque d'exposition si les contrôles d'accès manquent de précision. Sans mesures de sécurité robustes, les agents d'IA pourraient devenir un vecteur d'attaque majeur, ce qui éroderait la confiance des clients et exposerait les entreprises à des risques juridiques.

Comprendre les agents d'IA et leurs implications en termes de sécurité

Qu'est-ce qu'un agent d'IA ?

Un agent d'IA est un système logiciel autonome qui utilise les grands modèles de langage (LLM), le machine learning et des API pour effectuer des tâches sans intervention humaine directe. Contrairement aux logiciels traditionnels, les agents d'IA peuvent interpréter des entrées en langage naturel et y répondre, analyser des données en temps réel et effectuer des actions au nom des utilisateurs. Ces capacités en font des outils puissants pour les entreprises, mais imposent également un changement en profondeur des approches en matière de sécurité des identités. Les frameworks traditionnels dans ce domaine ont été conçus pour des utilisateurs humains, et non pour des logiciels d'IA autonomes capables de prendre des décisions indépendantes. Face à l'adoption des agents d'IA à grande échelle, les entreprises doivent repenser la gestion des identités et les contrôles d'accès pour optimiser la sécurité sans pour autant compromettre l'expérience utilisateur.

Pourquoi les entreprises adoptent des agents d'IA

Les entreprises, tous secteurs confondus, intègrent des agents d'IA pour améliorer l'efficacité, réduire les coûts et optimiser les expériences clients. Des chatbots d'IA qui gèrent les demandes de service client, aux agents de prise de décisions qui analysent les données métier, ces agents rationalisent les opérations et offrent des capacités inédites.

Des agents d'IA sont d'ores et déjà déployés dans de nombreux domaines pour optimiser les workflows, et ces cas d'usage vont se multiplier à mesure qu'un nombre croissant d'entreprises se feront à l'idée d'utiliser l'IA. Parmi les cas d'usage courants :

- **Automatisation du support client —** Les agents d'IA offrent un support de première ligne. Ils répondent aux questions courantes, résolvent les problèmes et font remonter les cas complexes à des agents humains. Ils contribuent ainsi à réduire les coûts et les délais de réponse, tout en préservant la satisfaction des clients.
- **Accélération des ventes et du marketing —** Les agents d'IA peuvent qualifier des leads, créer des ébauches de messages personnalisés et faciliter l'exécution de campagnes en fonction de données CRM en temps réel. Ils aident les équipes à gagner en productivité sans avoir à recruter de nouveaux talents.
- **Productivité interne et support IT —** Au sein des entreprises, les agents d'IA générative permettent aux collaborateurs de réinitialiser leurs mots de passe, de demander des accès, d'obtenir des réponses à leurs questions relatives aux RH ou aux politiques, et même de corriger du code, ce qui décharge de ces tâches de routine les équipes responsables de l'IT ou des opérations.

L'autonomie croissante des agents d'IA présente cependant un défi de sécurité : les agents non humains requièrent désormais des processus d'authentification, d'autorisation et de gouvernance. Avec l'adoption généralisée de l'IA, diverses parties prenantes des entreprises manifestent des préoccupations différentes concernant les risques de sécurité :

- **Les directeurs techniques** s'efforcent de trouver le juste compromis entre l'innovation dans le domaine de l'IA et la sécurité, en s'assurant que les agents d'IA sont intégrés en toute sécurité aux infrastructures existantes sans créer de vulnérabilités.
- **Les directeurs produits** s'attachent au maintien d'expériences clients fluides, tout en s'assurant que les interactions assistées par l'IA sont sécurisées et conformes.
- **Les directeurs des systèmes d'information (DSI)** supervisent la conformité réglementaire, la gestion des risques et la gouvernance des données afin de s'assurer que les agents d'IA respectent les standards sectoriels et les bonnes pratiques de sécurité.

Les défis de sécurité posés par les agents d'IA

Qu'est-ce que la RAG ?

La génération augmentée de récupération (RAG, Retrieval Augmented Generation) est une méthode d'IA qui améliore les réponses en commençant par rechercher des informations pertinentes dans un ensemble de documents, puis en utilisant ces données pour générer une réponse plus précise.

Pourquoi la mise en coffre des tokens est-elle importante ?

Auth for GenAI conserve les tokens utilisateurs dans un coffre (vault) sécurisé et gère le stockage, l'actualisation et l'accès afin que les développeurs n'aient pas à créer leur propre système de tokens.

Les systèmes d'IA et les applications assistées par l'IA sont complexes et exposés à un éventail de risques. En général, une vulnérabilité présente dans un système d'IA affecte également les applications assistées par l'IA qui en dépendent. Lors du développement d'agents d'IA, il faut tenir compte d'un certain nombre de défis de sécurité :

1. Premièrement, l'**authentification des utilisateurs**. L'agent ou l'application doit savoir qui est l'utilisateur. Par exemple, un chatbot peut avoir besoin d'afficher l'historique de chat, l'âge et le pays de résidence des utilisateurs pour personnaliser ses réponses. Cela exige une forme ou l'autre d'identification, qui peut être offerte par l'authentification.
2. Deuxièmement, l'**appel d'API au nom des utilisateurs**. Les agents d'IA se connectent à bien plus d'applications qu'une application web classique. À mesure que les applications d'IA générative intègrent un nombre croissant de produits, l'appel et le stockage sécurisé d'API seront essentiels.
3. Troisièmement, les **workflows asynchrones**. Les agents d'IA pourraient avoir besoin de plus de temps pour effectuer des tâches ou attendre que des conditions complexes soient remplies. Le délai pourrait se compter en minutes ou en heures, voire en jours. Les utilisateurs ne patienteront pas aussi longtemps. Ces cas vont devenir monnaie courante et seront implémentés sous forme de workflows asynchrones, avec des agents s'exécutant en arrière-plan. Ces scénarios seront supervisés par des humains, qui approuveront ou rejetteront des actions en dehors des chatbots.
4. Quatrièmement, l'autorisation pour la **génération augmentée de récupération (RAG, Retrieval Augmented Generation).** Presque toutes les applications d'IA générative peuvent alimenter des modèles d'IA avec des informations issues de plusieurs systèmes en vue d'implémenter la RAG. Pour éviter la divulgation d'informations sensibles, toutes les données alimentant des modèles d'IA pour répondre ou agir au nom d'un utilisateur doivent être des données auxquelles l'utilisateur a l'autorisation d'accéder.

Pour exploiter le plein potentiel de l'IA générative, nous devons relever ces quatre défis. Que vous développiez votre propre framework d'IA générative sur mesure dans un langage tel que Python, ou que vous utilisiez l'un des nombreux frameworks qui ont vu le jour ces deux dernières années, ces exigences doivent être satisfaites.

Il n'existe aucun plan d'action établi pour intégrer l'IA aux applications en toute sécurité. Les développeurs travaillant dans des entreprises élaborent des solutions maison pour créer l'agent d'IA en lui-même.

Protéger vos applications d'IA générative des menaces ciblant l'identité

C'est la raison pour laquelle nous avons développé Auth for GenAI. Auth for GenAI s'appuie sur tous les enseignements que nous avons tirés de notre travail avec des frameworks d'IA générative et des développeurs produits, ainsi que sur les dix ans d'expérience d'Auth0 dans le domaine de l'identité.

Parmi les avantages d'Auth for GenAI :

Authentification pour l'IA générative

Implémentez une expérience de connexion sur mesure pour les agents d'IA. Cela inclut la liaison de tous les comptes pour le profil utilisateur et l'authentification renforcée.

Mise en coffre des tokens

Utilisez des standards sécurisés pour connecter des agents d'IA à des outils tels que Gmail et Slack à l'aide d'OAuth 2.0 pour la gestion des tokens, tout en traitant automatiquement les actualisations et les échanges de tokens.

Autorisation asynchrone

Permettez aux agents d'IA d'effectuer des tâches avec des approbations « humain dans la boucle » (human-in-the-loop).

Autorisation granulaire pour la RAG

Protégez les données sensibles en autorisant les agents d'IA à ne récupérer que les documents auxquels l'utilisateur a accès.

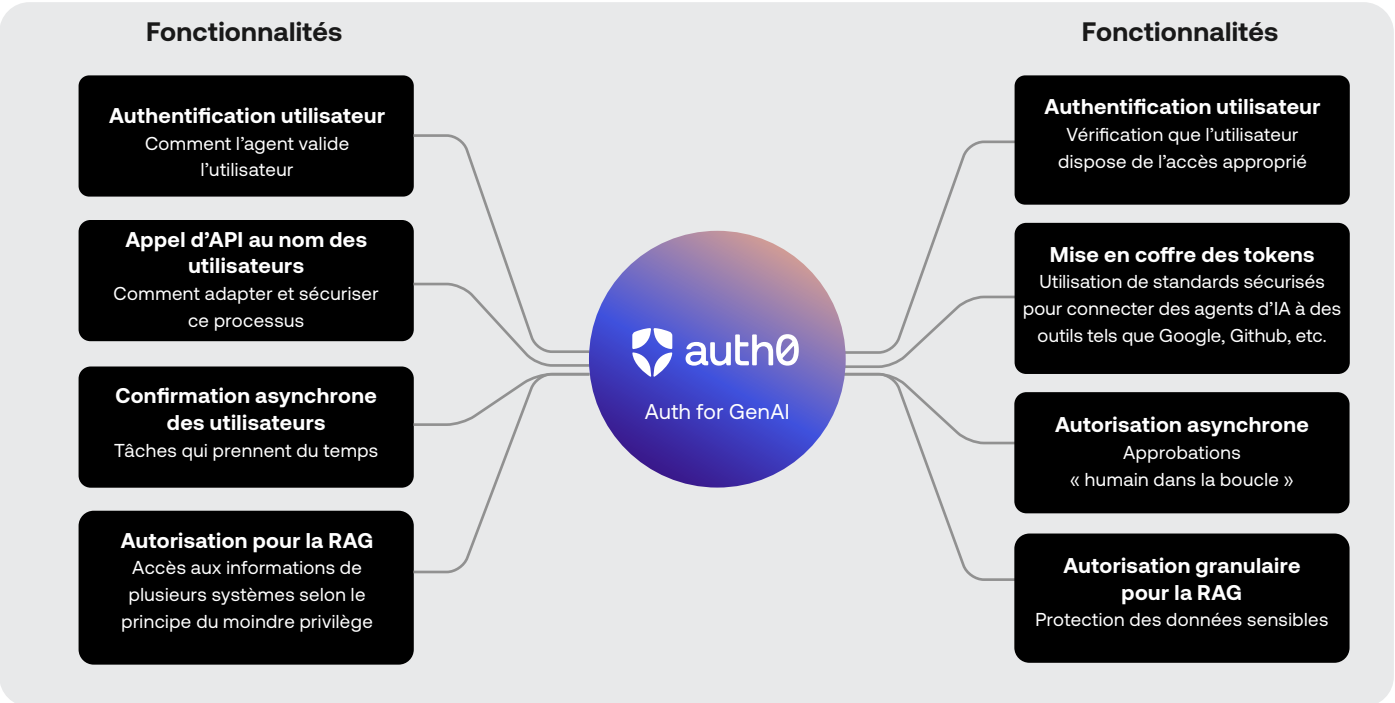
Auth for GenAI permet aux développeurs d'obtenir les mêmes résultats avec quelques lignes de code seulement, plutôt que de passer d'innombrables heures à coder afin de s'assurer que leurs applications d'IA sont sécurisées.

Une voie stratégique à suivre

Avec l'accélération de l'adoption de l'IA, les responsables de la sécurité doivent atténuer de façon proactive les risques pour l'identité associés aux agents d'IA. Les entreprises doivent :

1. **Évaluer les lacunes existantes dans la sécurité de l'IA** en procédant à un audit des politiques d'accès et des mécanismes d'authentification en place
2. **Implémenter des solutions IAM propres à l'IA** pour renforcer l'authentification, l'autorisation et les contrôles de surveillance
3. **Adopter une posture de sécurité continue** qui inclut une détection des anomalies en temps réel et des mécanismes de réponse automatisés
4. **Assurer la conformité réglementaire** en alignant les workflows des agents d'IA sur les standards sectoriels et les politiques de gouvernance
5. **Investir dans une sécurité des identités pérenne** qui s'adapte au paysage en constante évolution de l'automatisation pilotée par l'IA

L'IA change la façon dont les humains interagissent entre eux et avec la technologie. Au cours des dix prochaines années, nous allons assister à l'essor d'un vaste écosystème d'agents d'IA — des réseaux de programmes d'IA interconnectés qui s'intègrent à nos applications et agissent de manière autonome à notre place. Bien que l'IA générative présente de nombreux avantages, elle s'accompagne également de risques de sécurité non négligeables qui doivent être pris en compte lors du développement d'applications d'IA. Il est primordial de permettre aux développeurs d'intégrer en toute sécurité l'IA générative à leurs applications afin de les rendre adaptées à l'IA et au monde des entreprises.



Par ailleurs, comme les agents d'IA générative commencent à être intégrés à un nombre croissant d'applications et de services, il devient essentiel de disposer d'un accès structuré au contexte utilisateur. De nouveaux standards tels que le protocole MCP (Model Context Protocol) d'Anthropic offrent aux agents d'IA un moyen sûr et standardisé de récupérer du contexte — par exemple des événements de calendrier, des e-mails ou des documents — tout en respectant la confidentialité et les autorisations.

Mais le partage de contexte n'est pas sans risques : si l'accès n'est pas correctement protégé par des contrôles d'identité et d'autorisation, les agents pourraient exposer des données sensibles ou agir de manière inappropriée. L'implémentation par Auth0 d'un serveur MCP favorise une gestion sécurisée des processus d'authentification et d'autorisation au sein de ce framework. Les développeurs d'applications d'IA doivent intégrer des contrôles d'accès granulaires à ces nouveaux flux de contexte afin que seules les bonnes informations soient partagées avec le bon agent pour la bonne tâche.

L'IA est une réalité, pour le meilleur et pour le pire. Elle offre d'innombrables avantages aux utilisateurs et aux développeurs, mais s'accompagne également de préoccupations et de défis inédits sur le plan de la sécurité.

Avec Auth0 Platform, Okta vous décharge du fardeau de l'identité. Pour en savoir plus sur le développement sécurisé d'applications d'IA générative, rendez-vous sur auth0.com/ai.

À propos d'Auth0
Auth0® propose une approche moderne de la gestion des identités et permet aux entreprises d'offrir à tous leurs utilisateurs un accès sécurisé à n'importe quelle application. Auth0 est un produit hautement personnalisable, facile d'utilisation et suffisamment flexible pour répondre à tous les besoins des équipes de développement. Protégeant des milliards de transactions de connexion chaque mois, Auth0 allie ergonomie, confidentialité et sécurité afin de permettre aux clients de se concentrer sur l'innovation. Auth0 fait partie d'Okta, Inc., The World's Identity Company™.