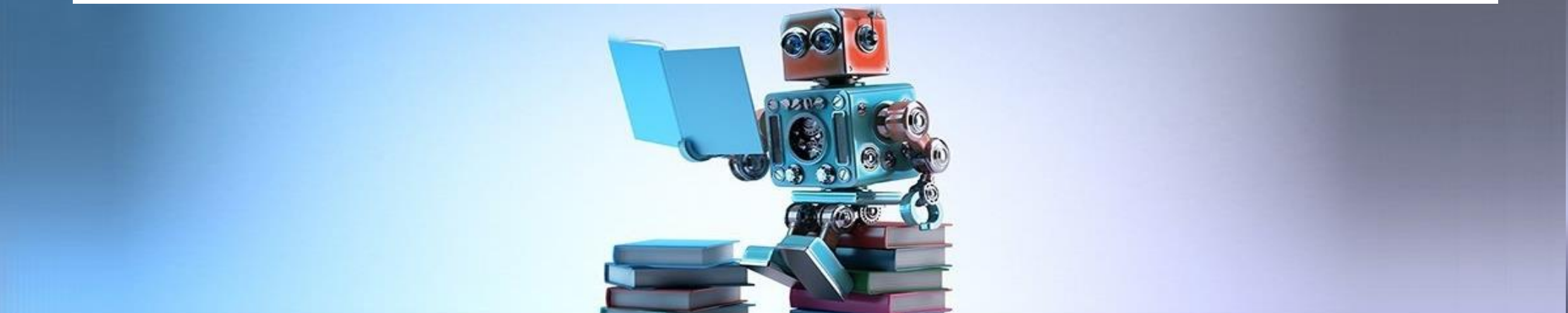
LeMagIT
L'informatique pour et par les pros
TechTarget
E-handbook
Contenu PRO+

Intelligence Artificielle et Cybersécurité : duo gagnant



Dans ce guide

Introduction.

Les techniques d'apprentissage automatique, et plus généralement d'intelligence artificielle, étaient les vedettes de l'édition 2015 de RSA Conference. Quelques années plus tard, elles se sont banalisées, s'invitant dans un grand nombre de systèmes de sécurité traditionnels, de la protection du poste de travail à la prévention des fuites de données en passant par la détection d'intrusion. Entre rachats et développements internes, elles sont désormais incontournables.

Dans les entreprises, les premiers bénéfices sont là, bien concrets et appréciés, au moins dans certains domaines. Mais ces techniques n'introduisent-elles pas de nouveaux défis ? La réponse dans ce dossier spécial.

Dans ce guide

Sécurité : est-on passé à trop d'intelligence artificielle ?

Valéry Marchive, journaliste

L'apprentissage automatique appliqué à l'analyse comportementale : cette technique, apparue commercialement en grande pompe il y a quatre ans, s'est aujourd'hui immiscée dans nombre de produits. Trop, peut-être ?

Exabeam fait partie de ces précurseurs de la détection d'anomalies comportementales appliquée à la sécurité informatique, la fameuse *UBA* (pour les utilisateurs) ou *UEBA* (pour les hôtes du SI en plus des utilisateurs). Sylvain Gil, son vice-président Exabeam en charge des produits, observe sereinement l'intégration de ce qu'il qualifie de « capacité technique » dans un nombre toujours croissant de systèmes de sécurité : « comme les règles de corrélation historiques, c'est quelque chose qui a vocation à être présent dans tous les contrôles de sécurité ».

Alors pour lui, il ne faut surtout pas hésiter à activer ces capacités d'analyse partout où elles sont disponibles : « allumez-les ! Le risque, ce n'est pas d'avoir plus de faux positifs, c'est de voir des choses que l'on n'aurait pas vues autrement ». La crainte de la menace qui passerait inaperçue du fait de cet étage de filtrage que certains qualifient volontiers d'intelligent ? « C'est un

Dans ce guide

problème qui existe depuis toujours ». Rien de nouveau sous le soleil, donc, ce ce côté-là, sinon, justement, une atténuation potentielle de ce risque. Alors pour Sylvain Gil, cela ne fait aucun doute : « au lieu des signatures, il vaut mieux essayer de comprendre quels sont les mauvais comportements, essayer de comprendre comment fonctionne l'environnement au jour le jour ».

Du fait de sa position, Sylvain Gil pourrait être soupçonné d'être partisan. Mais il n'est pas le seul à afficher ce point de vue. Son regard renvoie ainsi à l'analyse de Piotr Matusiak, RSSI de L'Oréal, qui relevait dans nos colonnes, à l'automne que « lorsque Darktrace signale quelque chose, ce n'est pas un faux positif, c'est une anomalie ! ». Et pour lui, celle-ci mérite investigation : « on y passe du temps, c'est sûr. Mais je préfère ce genre d'analyse plutôt que de me contenter de développer des scénarios statiques ».

Nicolas Fernandez et Paul Lemesle, respectivement directeur de la cybersécurité de Saint Gobain et responsable de son SOC, témoignaient, lors des Assises de la Sécurité, en octobre dernier, à Monaco, de leur déploiement des outils de Vectra, un concurrent direct de Darktrace, lui aussi spécialiste de l'analyse comportementale réseau. A cette occasion, ils relativisaient certains discours : « on parle d'intelligence artificielle, mais dans la réalité, c'est plutôt de l'apprentissage automatique ». Pour l'heure, la technologie vise à aider les analystes. Un choix stratégique à l'issue d'une situation – l'incident NotPetya – appréhendé comme une chance : « nous sommes dans une situation où nous sommes obligés de reconstruire ».

Dans ce guide

Mais la détection d'anomalie doit-elle être activée partout dans tous les contrôles de sécurité où elle est disponible… ou seulement sur un entrepôt de journaux centralisé ? « Il y a de vraies difficultés techniques » à cette seconde approche, relèvent-ils.

Emmanuel Gras, Pdg et co-fondateur d'Alsid, ne les contredira pas. Si oui, il a déjà pu observer des approches consistant à consolider de vastes volumes de détails d'activité dans de « grands datalakes », pour « essayer de mettre de l'intelligence au-dessus », c'est surtout réservé à une frange haute du marché : « des grandes banques américaines », par exemple. Mais dans la pratique, à ce jour… « je n'ai jamais vu cette stratégie fonctionner à l'échelle, ne serait-ce que compte tenu de la volumétrie ».

Dès lors, pour Emmanuel Gras, « il faut s'appuyer sur une solution pouvant apporter un premier niveau d'intelligence et d'analyse avant de remonter l'information un niveau au-dessus. Ce n'est pas un filtrage par périmètre, mais par fonction de sécurité ».

Pour autant, le filtrage par périmètre est également parfois retenu. Toujours lors des Assises de la Sécurité, Sopra Steria annonçait une offre centrée sur la sécurité des données, développée en partenariat avec Forcepoint et Brainwave. L'occasion d'un échange avec Fabien Lecoq, directeur sécurité technique au sein de l'entreprise de services numériques (ESN). Car des algorithmes d'apprentissage automatique sont mis à contribution, avec l'aide d'analystes, pour établir les clusters de classification de fichiers.

Dans ce guide

Et une grande banque cliente de l'ESN a fait le choix de l'UEBA, mais « sur 10 % des utilisateurs, en se concentrant sur les fonctions métiers où le risque est le plus élevé ». Du coup, « on ne s'attend pas à une généralisation de l'UEBA à 100 % sur une entreprise ».

Dans ce guide

Balazs Scheidler, One Identity : « considérer l'UBA comme un segment de marché était une erreur »

Valéry Marchive, journaliste

L'ancien directeur technique de Balabit, racheté par One Identity début 2018, se penche sur la manière dont l'analyse comportementale s'impose dans un éventail croissant de solutions de sécurité.

Il y a quatre ans, l'analyse comportementale appliquée à la sécurité était sur toutes les lèvres. Un nouvel eldorado avec ses porte-étendards, dont Fortscale, finaliste de l'Innovation Sandbox de l'édition 2015 de RSA Conference. Le marché était encombré par une multitude d'acteurs misant tantôt sur l'analyse des journaux d'activité, sur celle des points de terminaison de l'infrastructure à partir d'agents résidents, ou encore sur le trafic réseau (NTA, Network Trafic Analytics). Mais ce temps semble désormais bien loin, et Gartner anticipe la disparition de l'analyse comportementale comme marché isolé à l'horizon 2021. A travers une série d'articles, nous vous proposons de découvrir le regard que portent plusieurs experts sur cette évolution.

Dans ce guide

Balazs Scheidler, One Identity : Je pense que c'était une erreur de considérer les outils d'analyse comportementale comme une catégorie de marché distincte, car l'apprentissage automatique et les autres technologies utilisées par ces outils ne sont que cela : une technologie. Et c'est une erreur que de catégoriser des produits sur la base de la technologie qu'ils emploient ; ils devraient être plutôt catégorisés suivant les problèmes qu'ils essaient de résoudre.

Dans cette perspective, le fait que l'analyse comportementale et l'apprentissage automatique gravitent dans d'autres domaines est une bonne tendance. L'apprentissage automatique peut être utile dans la gestion des informations et des événements de sécurité (SIEM), dans la détection de menaces sur les hôtes de l'infrastructure (EDR) ou la gestion des comptes à privilèges (PAM). Les mêmes mathématiques peuvent être utilisées, mais puisque que les problèmes auxquels elles sont appliquées sont différents, les bénéfices le sont également.

Cela étant dit, je ne pense pas que ces technologies soient redondantes : au contraire, elles contribuent à fournir un meilleur service. Et en les implémentant dans des produits plus spécifiques, elles peuvent commencer à produire des résultats mesurables.

Par exemple, l'apprentissage automatique peut être utilisé pour détecter des anomalies dans des données de séries temporelles basées sur des comportements passés. Cela peut être utile dans un SIEM. L'apprentissage

Dans ce guide

automatique peut aussi être utilisé pour découvrir des groupes d'entités, et les produits de gestion des identités et des accès (IAM) ont commencé à profiter de fonctionnalités de découverte automatique de groupes et de rôles. Et puis l'apprentissage automatique peut être utilisé pour trouver des activités anormales, pour permettre notamment la découverte de détournement de comptes à privilèges.

Dans ce guide

Vincent Dely, Digital Guardian : « l'UEBA aide l'analyste à prioriser ses actions »

Valéry Marchive Bardy, journaliste

L'architecte solutions EMEA de Digital Guardian, se penche sur la manière dont l'analyse comportementale s'impose dans un éventail croissant de solutions de sécurité.

Il y a quatre ans, l'analyse comportementale appliquée à la sécurité était sur toutes les lèvres. Un nouvel eldorado avec ses porte-étendards, dont Fortscale, finaliste de l'Innovation Sandbox de l'édition 2015 de RSA Conference. Le marché était encombré par une multitude d'acteurs misant tantôt sur l'analyse des journaux d'activité, sur celle des points de terminaison de l'infrastructure à partir d'agents résidents, ou encore sur le trafic réseau (NTA, Network Trafic Analytics). Mais ce temps semble désormais bien loin, et Gartner anticipe la disparition de l'analyse comportementale comme marché isolé à l'horizon 2021. De fait, elle est de plus en plus présente dans les systèmes de gestion des informations et événements de sécurité (SIEM), les passerelles d'accès cloud sécurisé (CASB), les systèmes de prévention des fuites de données (DLP), ceux de gestion des accès et des identités (IAM) ou encore des comptes à

Dans ce guide

privilèges (PAM). A travers une série d'articles, nous vous proposons de découvrir le regard que portent plusieurs experts sur cette évolution.

Vincent Dely, Digital Guardian : L'UEBA est effectivement à la 'mode' pour ces différents outils. Au-delà de l'aspect marketing, cette technologie tente aussi de répondre à une demande des clients. Ils sont tous submergés d'informations à traiter pour délivrer un niveau suffisant de protection de leurs infrastructures et de leurs données sensibles. En parallèle, ils manquent de ressources aussi bien en nombre de personnes en charge d'analyser ces données, qu'en compétences pour en tirer une analyse efficace et approfondie.

Donc, la technologie d'UEBA tente d'aider à prioriser les opérations d'analyse en attirant l'attention sur les comportements ou les événements qui paraissent les plus suspects. Il s'agit davantage d'un point d'entrée dans le processus d'analyse cyber ou DLP que d'une finalité. De plus, cette technologie, en tentant d'identifier les modifications de comportements – pas uniquement des utilisateurs mais aussi des applications, etc. – au plus tôt, peut permettre de réagir plus en amont sur l'apparition d'une situation à risque. Plutôt que de réagir une fois l'incident terminé, on peut envisager d'agir dès les premières étapes et ainsi réduire la gravité de l'incident.

J'en conviens, la technologie d'UEBA n'est pas nécessairement adaptée à toutes les technologies que vous citez. Néanmoins, dans le cas de Digital Guardian, qui se trouve au plus près de la donnée, de l'utilisateur et des actions

Dans ce guide

qu'il entreprend sur ces données, la notion d'observation du comportement et de ses variations prend un sens naturel dans l'évolution de notre outil.

Cela pourrait être comparable à une salle de contrôle vidéo pour une commune. Un petit groupe d'agents ne pouvant suivre toutes les caméras en même temps, il est nécessaire que des systèmes automatiques attirent leur attention sur les événements importants. Mais cela n'enlève en rien la nécessité d'une interprétation humaine de la situation et de sa gravité.

Il est encore un peu tôt pour tirer des conclusions sur l'avenir de ce type d'approche, mais il est clair que les attaques sont de plus en plus difficiles à détecter depuis le réseau et, si l'on exclut les attaques en déni de service, toutes les autres attaques s'appuient sur des postes de travail ou des serveurs pour se réaliser. C'est donc à cet endroit qu'il faut se trouver pour les identifier et les contrer au mieux. C'est dans ce contexte que Digital Guardian se positionne et développe son offre. Notre récent classement comme leader dans le monde de l'EDR par Forrester confirme cette tendance.

L'équipe dirigeante de Digital Guardian ayant également créé Nitro Security, elle affiche la volonté de faire converger les capacités historiques de l'agent Digital Guardian à voir les activités et identifier les données sur les postes avec les technologies avancées d'analyse et de reporting dans le but de délivrer une solution unique pour la protection des postes de travail.

Dans ce guide

Il y a donc pour moi une complémentarité avec les approches basées sur le réseau comme le SIEM et les systèmes de protection des infrastructures comme les firewall NG. Digital Guardian se charge d'analyser et de comprendre ce qui se passe en détail au niveau de chaque poste de travail et peut remonter vers un composant central de pilotage du SOC, comme le SIEM, les alertes les plus pertinentes. Il est alors possible pour les opérateurs d'approfondir l'analyse en s'appuyant sur le Analytics and Reporting Cloud (ARC) de Digital Guardian pour collecter les éléments de contexte détaillés.

Nous en sommes à la version 2.0 de notre moteur UEBA et nous prévoyons de faire évoluer la plateforme pour permettre aux clients d'intégrer les définitions des comportements qu'ils souhaitent voir analysées dans le modèle pour rendre le système plus pertinent aux particularités de leur environnement. Il est également prévu d'intégrer un organigramme de l'entreprise pour estimer l'impact d'un incident isolé à l'échelle de la hiérarchie. Un utilisateur isolé pouvant mettre à risque la direction générale de l'entreprise.

Dans ce guide

Christian Have, LogPoint : « chaque capacité d'UEBA ajoute du contexte dans le SIEM »

Valéry Marchive, journaliste

Le chef de produits de l'éditeur, qui a lui-même développé son propre moteur d'analyse comportementale, se penche sur la manière dont cette technologie s'impose dans un éventail croissant de solutions de sécurité.

Il y a quatre ans, l'analyse comportementale appliquée à la sécurité était sur toutes les lèvres. Un nouvel eldorado avec ses porte-étendards, dont Fortscale, finaliste de l'Innovation Sandbox de l'édition 2015 de RSA Conference. Le marché était encombré par une multitude d'acteurs misant tantôt sur l'analyse des journaux d'activité, tantôt sur celle des points de terminaison de l'infrastructure à partir d'agents résidents, tantôt sur le trafic réseau (NTA, Network Trafic Analytics). Mais ce temps semble désormais bien loin. Gartner anticipe la disparition de l'analyse comportementale comme marché isolé à l'horizon 2021. A travers une série d'articles, nous vous proposons de découvrir le regard que portent plusieurs experts sur cette évolution.

Dans ce guide

LeMagIT : L'intérêt à intégrer un moteur d'analyse comportementale à un système de gestion des informations et des événements de sécurité apparaît évident. Mais l'analyse comportementale s'invite partout, de la détection des menaces sur les hôtes (EDR) jusqu'aux passerelles d'accès Cloud sécurité (CASB), en passant par la gestion des accès et des identités (IAM) et la prévention des fuites de données (DLP). Ce n'est pas un peu trop ?

Christian Have : L'UEBA est une capacité fantastique qui a toute sa place dans toutes sortes de technologies. C'est pourquoi l'avoir dans de nombreux domaines et en particulier dans le SIEM est pertinent : grâce à la présence de l'UEBA dans des outils spécifiques, vous pouvez identifier, avec un haut degré de connaissance du domaine, exactement ce qui est anormal et ce qui ne l'est pas. La présence de l'UEBA dans l'EDR et le CASB permet à ces systèmes de bloquer et d'alerter sur les observations. Les SIEM sont le dernier kilomètre de l'analyse.

Les alertes provenant de capacités d'UEBA hors du SIEM ne font qu'ajouter du contexte à ce qui est observé dans le SIEM ; ce n'est certainement pas l'un ou l'autre !

LeMagIT : N'y a-t-il pas un risque de passer à côté de quelque chose si l'on enchaîne les moteurs d'analyse comportementale appliqués à différents sous-ensembles de données et que l'on se contente de s'intéresser aux alertes qu'ils génèrent ?

Dans ce guide

Christian Have : Avoir plus de faux positifs est, bien sûr, un problème. Mais les moteurs ne sont pas enchaînés. Notre moteur d'analyse comportementale examine peut-être 50 ou 100 plate-formes et applications différentes au sein d'un réseau client, et c'est le total qui donne de la lumière aux indications.

LeMagIT : En fin de compte, que recommanderiez-vous comme mise en œuvre ?

Christian Have : Les capacités d'analyse comportementale présentes dans différents outils spécifiques ont l'avantage de permettre de bloquer automatiquement des menaces, comme dans un système de prévention d'intrusion (IPS), par opposition à un système de détection d'intrusion (IDS). L'UEBA intégrée au SIEM est ce qui permet à l'analyste de hiérarchiser ce qui est observé dans les différentes sources de données avec plus de précision et de rapidité. Il est donc certainement bénéfique pour l'analyste de comprendre l'étendue totale des menaces bloquées par des capacités d'UEBA locales, et des menaces observées par l'UEBA piloté par le SIEM.

Dans ce guide

Dave Palmer, Darktrace : « la qualité des données utilisées pour la modélisation est essentielle »

Valéry Marchive, journaliste

Le directeur technique de ce spécialiste des techniques d'intelligence artificielle appliquées à l'analyse du trafic réseau se penche sur les enjeux et les perspectives de cette nouvelle approche.

Il y a quatre ans, l'analyse comportementale appliquée à la sécurité était sur toutes les lèvres. Un nouvel eldorado avec ses porte-étendards, dont Fortscale, finaliste de l'Innovation Sandbox de l'édition 2015 de RSA Conference. Le marché était encombré par une multitude d'acteurs misant tantôt sur l'analyse des journaux d'activité, sur celle des points de terminaison de l'infrastructure à partir d'agents résidents, ou encore sur le trafic réseau (NTA, Network Trafic Analytics). Mais ce temps semble désormais bien loin, et Gartner anticipe la disparition de l'analyse comportementale comme marché isolé à l'horizon 2021. A travers une série d'articles, nous vous proposons de découvrir le regard que portent plusieurs experts sur cette évolution.

Dans ce guide

Dave Palmer, Darktrace : Nous vivons une époque où AI, ou ML, deviennent la norme pour programmer, développer des produits. Il ne s'agit pas de remplacer la programmation linéaire à laquelle nous sommes habitués, mais devenir une partie normale de la manière dont on interagit avec les ordinateurs, en particulier lorsque l'on traite de problèmes liés à de grand volumes de données, comme la sécurité.

Mais une question à poser consiste à savoir si l'apprentissage automatique est utilisé pour simplifier la vie de l'utilisateur ou celle du développeur. Et justement, dans de nombreux cas en sécurité, aujourd'hui, il ne s'agit pas d'améliorer la vie du client mais celle des personnes qui font le produit.

Par exemple, dans beaucoup de produits de protection des hôtes, postes de travail et serveurs, ou de sécurité réseau, l'intelligence artificielle et l'apprentissage automatique sont surtout là pour réduire les efforts nécessaires pour faire le produit.

Ce que je veux dire par là, c'est : « est-ce que les nouveaux anti-virus sont fondamentalement différents des produits historiques, basés sur des heuristiques ? Je ne le pense pas. Je pense qu'ils sont essentiellement construits d'une autre manière, d'une façon plus appropriée pour faire grandir votre activité en tant qu'éditeur ».

Le cas des produits intégrant des capacités de modélisation du comportement des utilisateurs est différent, que l'on parle de système de prévention des fuites

Dans ce guide

de données (DLP), de passerelle d'accès Cloud sécurisé (CASB), de système d'analyse du trafic réseau (NTA), voire même de système de gestion des informations et des événements de sécurité (SIEM).

Là, je pense qu'il faut être prudent dans le choix d'activités ou non des capacités d'analyse comportementale, car la qualité des données utilisées pour la modélisation varie selon les domaines. Ainsi, si vous avez un déploiement de SIEM acceptable, mais sans couvrir toute l'infrastructure, et un système d'UEBA qui ne peut pas voir des choses très importantes comme ce qui se passe dans les applications cloud, vous risquez de buter rapidement sur l'effet « sapin de Noël » : l'incomplétude de la vision risque de conduire à de mauvaises décisions et alertes.

Si les données utilisées en entrée ne sont pas bonnes, on risque d'aggraver la situation. C'est pour cela que l'on s'attache à appliquer ces technologies au cœur du réseau, et au cœur du cloud, où il est plus facile de s'assurer que les bonnes données viennent alimenter le système.

Les SIEM et les moteurs d'UEBA autonomes dépendent fortement de décisions humaines sur les sources de données. Et il est malheureusement très facile d'exclure du périmètre des données dont un responsable sécurité sera tenté de penser qu'elles ne comptent pas… alors même qu'elles peuvent être essentielles pour l'application de techniques d'intelligence artificielle et d'apprentissage automatique. Tout simplement parce qu'elles apportent du contexte sur ce qui se passe réellement.

Dans ce guide

LeMagIT : Suggérez-vous que l'on demande à l'intelligence artificielle, dans un SIEM par exemple, de conseiller sur le type de données à surveiller ?

Dave Palmer, Darktrace : Je pense que c'est inévitable. Aujourd'hui, c'est un humain qui décide des données devant être collectées. Mais exclure des données, c'est prendre le risque de se priver d'éléments de contexte, tout aussi utiles à des algorithmes d'apprentissage automatique qu'à des analystes humains.

Et généralement, on n'envoie pas toutes les données dans un SIEM, seulement les alertes ou les événements suspects, ce qui peut indiquer une attaque. C'est pourquoi l'apprentissage automatique peut être là mis en difficulté.

Donc, l'avenir est clairement à la mise en place d'assistant à intelligence artificielle, peut-être sur chaque poste de travail ou serveur, capable de prendre de meilleures décisions quant aux données à transmettre au SIEM. Mais aussi peut-être à des approches plus interactives, où la base de données entre les deux demande spontanément plus d'informations autour de tel ou tel événement.

Je pense que c'est la seule manière dont l'intelligence artificielle et l'apprentissage automatique permettront une évolution des approches traditionnelles.

Dans ce guide

Mais on peut imaginer également d'autres modèles, plus distribués, avec des groupes de machines discutant entre elles et traitant les données directement alors qu'elles transitent au sein du groupe.

Dans ce guide

Accéder à plus de contenu exclusif PRO+

Vous avez accès à cet e-Handbook en tant que membre via notre offre PRO+ : une collection de publications gratuites et offres spéciales rassemblées pour vous par nos partenaires et sur tout notre réseau de sites internet.

L'offre PRO+ est gratuite et réservée aux membres du réseau de sites internet TechTarget.

Profitez de tous les avantages liés à votre abonnement sur: http://www.lemagit.fr/eproducts

Images; Fotolia

Dans ce guide

Cyrille Chausson | *Rédacteur en Chef*
TechTarget
22 rue Léon Jouhaux, 75010 Paris
www.techtarget.com