N°233 — Février 2025

www.linformaticien.com

L'INFORMATICIEN

RH
Préparer un audit de conformité

Étude
L'IA et sa consommation énergétique

Retex
L'IA au service de l'Hôpital

Hardware
Retour du CES

DOSSIER

Conformité

Quelles priorités pour 2025 ?

L 14614 - 233 - F: 8,50 € - RD

Les grandes ambitions produisent de grands résultats. Nos puissantes solutions IA peuvent vous aider à surmonter vos limites.

Rendez-vous sur **HPE.com/fr**

Hewlett Packard Enterprise

www.linformaticien.com
L'INFORMATICIEN

RÉDACTION

88 boulevard de la Villette, 75019 Paris, France.
Tél. : +33 (0)1 74 70 16 30 — *contact@linformaticien.com*

RÉDACTION : Bertrand Garé (rédacteur en chef)
et Victor Miget (rédacteur en chef adjoint)
avec : Patrick Brebion, Vincent Bussière, Christine Calais, Jérôme Cartegini, Alain Clapaud, Michel Chotard, François Cointe, Guillaume Renouard et Thierry Thaureaux

SECRÉTAIRE DE RÉDACTION : Amélie Ermenault Martin

MAQUETTE ET RÉALISATION : Franck Soulier (chef de studio)

PUBLICITÉ
Antoine Foulon — *afoulon@linformaticien.com*

VENTE AU NUMÉRO

France métropolitaine 8,50 € TTC (TVA 5,5 %)

ABONNEMENTS

France métropolitaine 72 € TTC (TVA 5,5 %)
magazine + numérique

Toutes les offres :
www.linformaticien.com/abonnement

Pour toute commande d'abonnement d'entreprise ou d'administration avec règlement par mandat administratif, adressez votre bon de commande à :

L'Informaticien, service abonnements,
88 boulevard de la Villette, 75019 Paris, France.
ou à abonnements@linformaticien.com

IMPRESSION

Imprimé en France par Imprimerie Chirat (42)
Dépôt légal : 1er trimestre 2025

L'INFORMATICIEN est publié par PC PRESSE, S. A. S.
au capital de 130 000 euros.
Siège social : 88 boulevard de la Villette, 75019 Paris, France.

ISSN 1637-5491

Une publication PCpresse,

FICADE

PRÉSIDENT, DIRECTEUR DE LA PUBLICATION :
Gaël Chervet

Il n'y pas que l'Europe qui régule !

Notre dossier traite de la cascade des conformités que devront respecter les entreprises dans les prochains jours, mois… Il est même difficile d'en faire le compte exhaustif tellement il y en a. Si, dans notre prisme, on insiste beaucoup sur celles prises au niveau européen comme NIS, DORA, l'IA Act, le Data Act… Les USA et la Chine ne sont pas en reste. On assiste à une véritable guerre des standards que se livrent les différentes puissances économiques de ce monde avec pour chacun leurs propres philosophies. Aux USA, la possibilité de faire de tout un business, en Europe de respecter les libertés fondamentales et l'individu. En Chine, renvoyer l'ascenseur sur les différentes mesures d'embargo décidées par les USA. Il est évident que la lutte autour des standards est tout aussi importante que les autres conflits, du fait de sa conséquence sur l'économie.

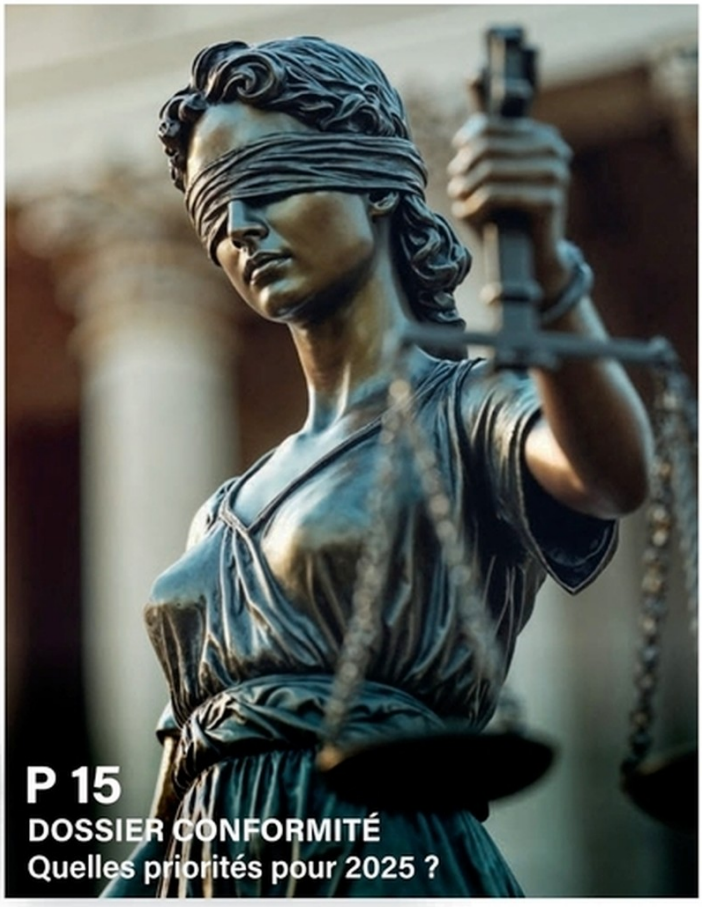
Dans notre supplément InfoCR, vous retrouverez ce qui va faire votre année avec une vision la plus large possible du paysage des attaques et des moyens de prévention à disposition. Il en est de même pour notre rubrique réseau qui compile les avis des principaux experts de Juniper sur les technologies du domaine pour 2025.

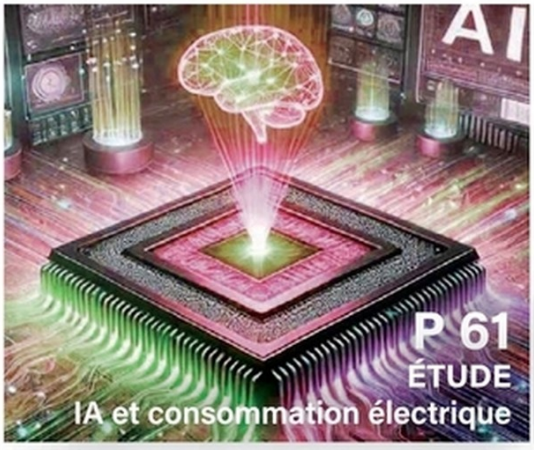
Autre sujet d'actualité autour de l'intelligence artificielle, avec un retour d'expérience menée au CHU de Montpellier pour alléger la charge de travail des personnels hospitaliers avec l'utilisation de différents outils intelligents.

Il en est encore temps, L'informaticien vous souhaite une superbe, joyeuse, enrichissante année 2025 ! ☐

Bertrand Garé
Rédacteur en Chef


WAICF
WORLD ARTIFICIAL INTELLIGENCE CANNES FESTIVAL
FEBRUARY
13-15, 2025
CANNES, FRANCE
WHERE AI CHANGE MAKERS
MEET INDUSTRY LEADERS
12 000 business attendees
250 exhibitors
320 speakers
www.worldaicannes.com
An event of
INSTITUT EUROPIA COMPRENDRE POUR AGIR
PALAIS DES FESTIVALS ET DES CONGRÈS CANNES
CANNES CÔTE D'AZUR FRANCE
DÉPARTEMENT DES ALPES-MARITIMES
Organized by
RX In the business of building businesses

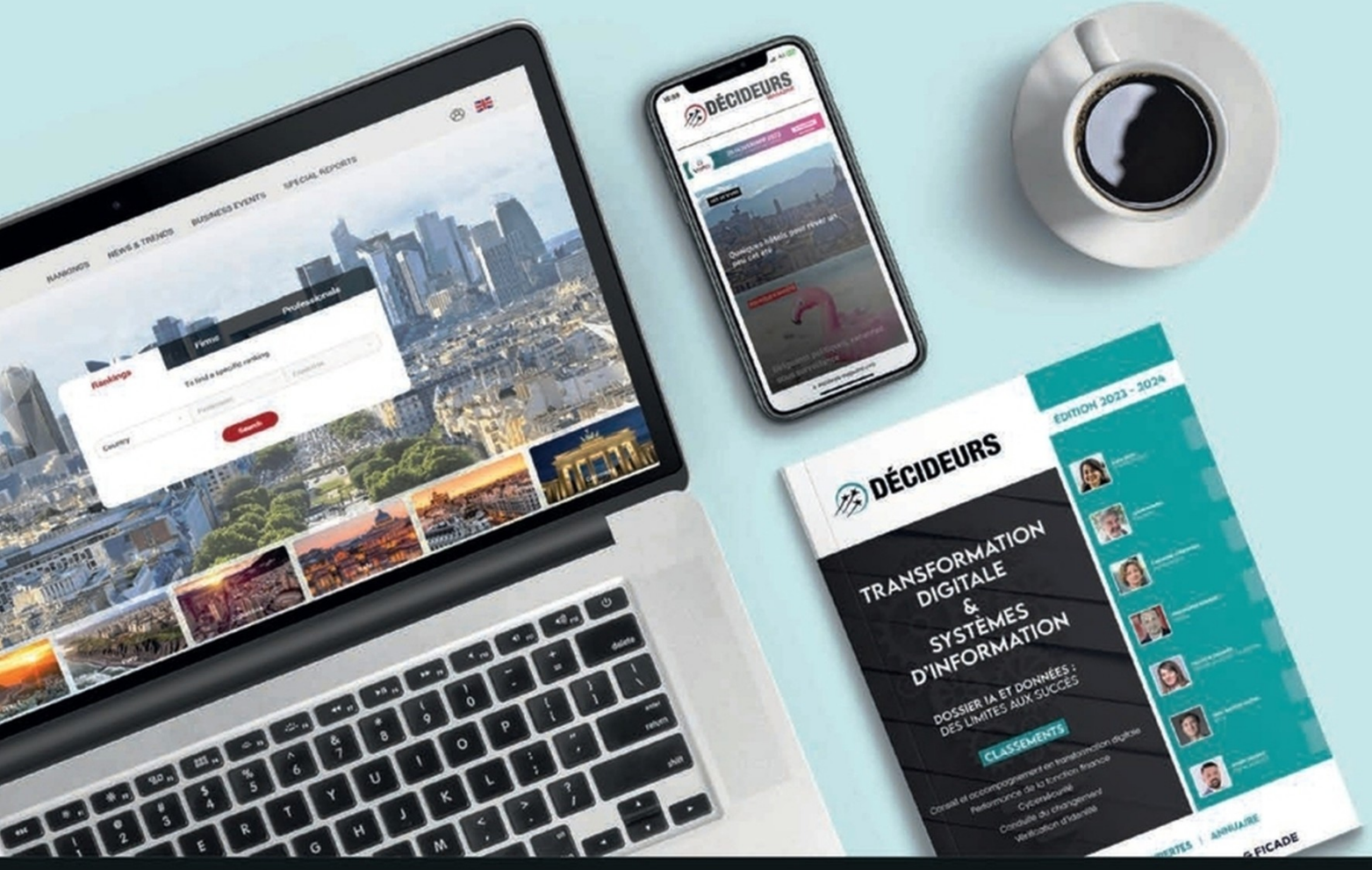




DÉCIDEURS

L'information stratégique pour **bien** choisir vos partenaires

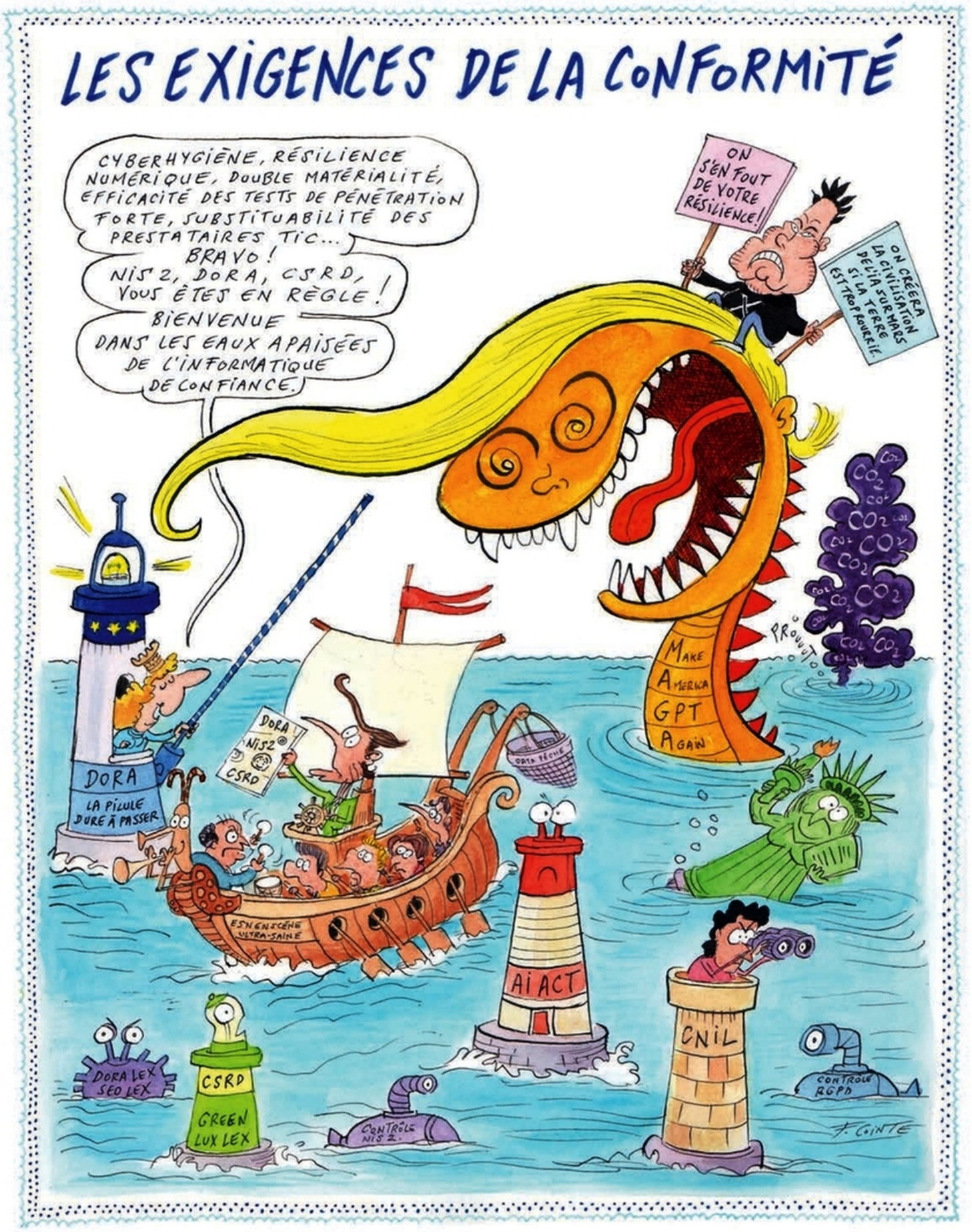
Conseil et accompagnement en transformation digitale | Performance de la fonction finance | Cybersécurité
Conduite du changement | Vérification d'identité



COMMANDER LE GUIDE
TRANSFORMATION DIGITALE & SYSTÈMES D'INFORMATION

DÉCIDEURS AGENCE DE NOTATION

FICADE


LES EXIGENCES DE LA CONFORMITÉ
CYBERHYGIÈNE, RÉSILIENCE NUMÉRIQUE, DOUBLE MATÉRIALITÉ, EFFICACITÉ DES TESTS DE PÉNÉTRATION FORTE, SUBSTITUABILITÉ DES PRESTATAIRES TIC...
BRAVO ! NIS 2, DORA, CSRD, VOUS ÊTES EN RÈGLE !
BIENVENUE DANS LES EAUX APAISÉES DE L'INFORMATIQUE DE CONFIANCE.
ON S'EN FOUT DE VOTRE RÉSILIENCE !
ON CRÉERA LA CIVILISATION SUR MARS DE L'IA SUR MARS SI LA TERRE EST TROP POURRIE.
MAKE AMERICA GPT AGAIN
PROUUUT
DORA
LA PILULE DURE À PASSER
DORA NIS2 CSRD
ESN EN SCÈNE ULTRA-SAINE
AI ACT
CNIL
DORA LEX SEO LEX
CSRD
GREEN LUX LEX
CONTRÔLE NIS 2.
CONTRÔLE RGPD
F. COINTE

FOCUS

Les États-Unis resserrent leur emprise sur les puces d'IA

L'administration Biden a présenté, lundi 13 janvier 2024, de nouvelles règles sur l'exportation de puces d'intelligence artificielle. Les États-Unis cherchent à empêcher leurs adversaires d'accéder aux dernières innovations tout en protégeant leur leadership. Si l'Oncle Sam dit ne pas vouloir empêcher les exportations vers ses alliés, l'Union européenne s'est toutefois émue de cette décision qui restreint l'accès aux exportations de puces d'IA pour certains États membres.

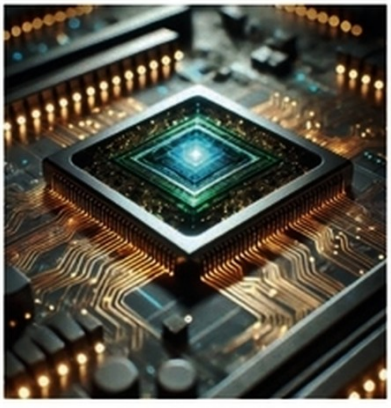
C'est l'ultime serrage de vis de l'administration Biden sur l'exportation de puces d'IA, une semaine avant l'investiture de Donald Trump. Lundi 13 janvier, la Maison-Blanche a annoncé de nouvelles restrictions à l'export des processeurs graphiques avancés (GPU), indispensables pour alimenter les centres de données nécessaires à l'entraînement des modèles d'IA.

Comme à son habitude, l'Oncle Sam a brandi la sécurité nationale et la nécessité de diffuser des technologies d'IA responsables et de confiance pour justifier ce nouveau serrage de vis à l'encontre de ceux qu'il qualifie « d'adversaires ». Washington dit craindre que le potentiel de l'IA, et plus particulièrement des technologies américaines les plus avancées, soit utilisé à des fins militaires, cybermalveillantes ou de surveillance de masse. Par ce biais, la première puissance mondiale tente aussi de préserver la confortable avance dont elle jouit dans le domaine des semi-conducteurs avancés, et veut faire en sorte que *« l'IA mondiale repose sur des bases américaines »*, peut-on lire dans un communiqué de presse de la Maison-Blanche.

Des exceptions (sous contrôle) pour les alliés

Dans les grandes lignes, Washington va imposer des restrictions d'exportation et de transfert de puces informatiques avancées pour une liste élargie de 120 pays, dont Israël, Singapour et l'Arabie saoudite. Les entreprises américaines devront obtenir des autorisations pour la vente dans ces pays. Les exportations seront purement et simplement bloquées pour les pays soumis à un embargo sur les armes, comme la Russie, la Chine et la Corée du Nord.

L'administration a également assuré que réaliser pleinement les bénéfices de



l'intelligence artificielle ne se ferait pas sans le concours des alliés et partenaires privés des États-Unis. C'est pourquoi des exceptions de licence ont été mises en place, autorisant l'exportation, la réexportation ou le transfert de puces informatiques avancées, sans autorisation préalable, vers ses 18 alliés et partenaires clés, dont la France, le Royaume-Uni, ou encore la Corée du Sud.

La Chine et l'Union européenne réagissent

La Chine a fustigé, dans un communiqué, ce qu'elle considère être un *« abus du contrôle des exportations et une violation flagrante des règles commerciales internationales »*. Le pays fait les frais, depuis plusieurs années déjà, d'une forte pression des États-Unis et de certains de ses alliés qui multiplient les restrictions d'exportations de puces avancées et de composants nécessaires à leur production. Une stratégie entamée par Donald Trump lors de sa première mandature, puis poursuivie par Joe Biden, qui avait mis en place d'importantes restrictions en octobre 2022 et octobre 2023, entre autres.

Malgré les déclarations de Washington vis-à-vis de ses alliés et partenaires les plus proches, la Commission européenne n'a pas non plus masqué son inquiétude. *« Nous sommes préoccupés par les mesures américaines adoptées aujourd'hui, restreignant l'accès aux exportations de puces d'IA avancées pour certains États membres de l'UE [comme la Lituanie et l'Estonie] et leurs entreprises »*, se sont émues la vice-présidente exécutive Henna Virkkunen et le commissaire Maroš Šefčovič, dans une déclaration conjointe. L'Union européenne estime en outre qu'il est *« dans l'intérêt économique et sécuritaire des États-Unis que l'UE achète des puces d'IA avancées aux États-Unis sans limitations »*.

C'est aussi l'avis de Nvidia, leader sur le segment des GPU, qui a dénoncé dans un communiqué une règle « mal conçue » qui *« risque de freiner la concurrence — moteur vital de l'innovation »*. L'entreprise américaine a ajouté que : *« comme l'a démontré la première administration Trump, l'Amérique triomphe grâce à l'innovation, à la concurrence et au partage de ses technologies avec le monde »*. Le texte doit entrer en vigueur dans 120 jours. Ce qui laisse le temps à la prochaine administration d'y mettre son empreinte. La Commission européenne a d'ores et déjà indiqué avoir exprimé ses *« préoccupations à l'administration américaine actuelle »*, avec qui elle s'est dite *« impatiente de collaborer de manière constructive »*.

Bien que cette collaboration s'annonce d'ores et déjà délicate, tant, au sein même de l'UE, tous les États membres ne semblent pas au diapason. Sur la même ligne que Washington, le gouvernement néerlandais a annoncé, mercredi 15 janvier, de nouvelles restrictions, et élargira dès le 1er avril un plus grand nombre d'équipements avancés de semi-conducteurs, qui seront soumis à une obligation d'autorisation nationale.

Microsoft et Amazon investissent des milliards dans leurs data centers

Après une année faste en 2024, les grands acteurs de l'IA démarrent fort et n'hésitent pas à mettre la main au portefeuille pour renforcer leurs positions respectives dans le domaine de l'intelligence artificielle. L'IA et le cloud ne seraient rien sans leur colonne vertébrale : l'infrastructure, indispensable pour exécuter les charges de travail liées à l'intelligence artificielle.
Et ça, les géants du secteur l'ont bien compris et financent leurs data centers à grands coups de dizaines de milliards.

C'est dans ce contexte que Microsoft a annoncé, par l'intermédiaire de son président, Brad Smith, un investissement de 80 milliards de dollars en 2025, pour construire des data centers dédiés à l'intelligence artificielle générative et aux applications basées sur l'IA et le cloud. Plus de la moitié de cette enveloppe sera dépensée aux États-Unis.
Cet investissement colossal doit soutenir l'innovation du secteur tant vantée par le président de Microsoft, et confirmer la confortable avance dont jouissent les États-Unis dans la course mondiale à l'intelligence artificielle. *« L'IA promet de stimuler l'innovation et d'accroître la productivité dans tous les secteurs de l'économie. Les États-Unis sont bien placés pour se maintenir à l'avant-garde de cette nouvelle vague technologique, à condition de tirer parti de leurs points forts et de s'associer efficacement à l'échelle internationale »*, a écrit Brad Smith dans un billet de blog.
Quelques jours après l'annonce de Microsoft, c'est AWS qui a annoncé qu'il injecterait 11 milliards de dollars (Md$) en Géorgie (États-Unis) pour son infrastructure de centres de données dédiée à l'intelligence artificielle et au cloud. Les centres de données exécuteront des GPU pour fournir de la puissance de calcul, afin d'exécuter les charges de travail liées au cloud et celles des modèles d'IA et de machine learning. AWS n'en est pas à son premier investissement dans cet état du sud des États-Unis. Depuis 2010, la filiale y a investi 18,5 Md$ et a contribué à hauteur de 20,1 Md$ du PIB de la Géorgie, affirme-t-elle.

IBM et GlobalFoundries tournent la page de leurs différends

On oublie tout et on se serre la pince. GF et IBM ont annoncé, jeudi 2 janvier 2025, avoir conclu un accord mettant fin « à toutes les affaires judiciaires, y compris les réclamations liées à des violations de contrat, des secrets commerciaux et des droits de propriété intellectuelle entre les deux parties », peut-on lire dans un court communiqué.

Petit rappel des faits : en 2014, IBM revend à GF son activité de fabrication de semi-conducteurs pour 1,5 milliard de dollars. Dans le cadre de cet accord, GF s'engage à produire des technologies 14 nanomètres (nm) pour les processeurs d'IBM et à développer des technologies à 10 nm et moins. Mais très vite, IBM constate des retards de production ainsi que l'abandon, dès 2018, des efforts de développement des processeurs 10 nm et 7 nm. IBM se tourne finalement vers Samsung pour se fournir en puces 7 nm, porte plainte en 2021 devant un tribunal new-yorkais contre GF, et exige le paiement de dommages et intérêts.
En 2023, c'est au tour du fondeur d'intenter une action en justice contre IBM,

qu'il accuse d'avoir divulgué des secrets commerciaux et des propriétés intellectuelles à ses partenaires Intel et au fabricant japonais de puces Rapidus.
Sans que l'on sache trop pourquoi, les deux entreprises ont finalement décidé d'enterrer la hache de guerre avec cet accord, dont les détails n'ont pas été révélés, qui *« ouvre la voie à de nouvelles opportunités de collaboration dans des domaines d'intérêt commun »*. Et des opportunités, il y en a. En effet, en mai 2021, IBM a réalisé une percée importante en développant la première puce au monde annoncée avec la technologie des nano-feuilles de 2 nm. En fin d'année dernière, Big Blue et Rapidus avaient, en outre, annoncé avoir relevé plusieurs défis techniques pour étendre la production de puces 2 nm à l'échelle industrielle. GF a, quant à lui, perçu une subvention de 1,5 milliard de dollars en 2024 pour renforcer ses capacités de fabrication, dans le cadre de la loi CHIPS and Science Act, qui vise à soutenir la production de semi-conducteurs aux États-Unis.

RACHATS

Onepoint avale Stack Labs

L'ESN, spécialisée dans le cloud et la donnée, élargit ses compétences en reprenant Stack Labs, une société opérant sur les mêmes domaines.
Fondée à Toulouse fin 2017 par Frédéric Volpi, Stack Labs est une société française spécialisée dans l'accompagnement des entreprises vers le cloud computing et la modernisation de leurs applications et plateformes de données. L'entreprise a su, dès ses débuts, lier des partenariats forts avec les clouds providers internationaux (Google et AWS) lui permettant d'être reconnue pour son expertise technique par ses clients à la fois grands comptes et ETI. Avec l'acquisition de Stack Labs, le groupe se positionne sur un marché en forte croissance (+30 %) et se dote ainsi de centres d'excellence technologique cloud situés à Toulouse et à Paris. De plus, One point étoffe sa capacité d'accompagnement des ETI et grands groupes, en France et à l'international.

Coralogix reprend Aporia

Coralogix met la main sur l'activité d'Aporia, un spécialiste de l'observabilité des processus IA. Aporia est une startup israélienne fournissant d'importants services pour l'intelligence artificielle comme l'observabilité des opérations, mais aussi le contrôle des règles d'utilisation et de garde-fous pour éviter les hallucinations et autres biais possibles. A l'occasion de ce rachat, coralogix lance un laboratoire spécialisé qui sera sous la houlette de deux dirigeants d'Aporia : le co-fondateur Liran Hason et le CTO Alon Gubkin.
Les technologies d'Aporia vont rejoindre Coralogix Services. L'idée est de fournir une plateforme unifiée pour l'observabilité de l'ensemble des opérations, d'IA et traditionnelles. Le montant de la transaction est estimé à 50 M$.

Datadog rachète Quickwit

L'éditeur de solution de monitoring des environnements cloud natifs reprend Quickwit, une solution de recherche dans les données open source, à destination des entreprises fortement régulées.
Le logiciel de Quickwit permet de réaliser des recherches sur les logs à une vaste échelle, afin de suivre l'utilisation, les coûts et les performances sur des volumes toujours en expansion. La solution peut se déployer sur site ou dans le Cloud.
Les capacités du logiciel Quickwit vont permettre à Datadog de gérer de nouveaux aspects sur les logs de ses clients : la localisation des données, et la conformité vis-à-vis de règles comme le RGPD ou d'autres règles associées aux données.
La première étape va être de conduire l'intégration du logiciel dans la plateforme de Datadog, alors que se profile une nouvelle version de Quickwit sous licence Apache V2.

Darktrace s'offre Cado Security

L'entreprise de cybersécurité Darktrace a acquis Cado Security, un fournisseur de solutions d'investigation et de réponse aux cybermenaces pour les environnements hybrides et multi-cloud.
Les outils de Cado Security couvrent une large gamme d'environnements, allant du multi-cloud aux conteneurs, en passant par le SaaS et les infrastructures on-premise. Darktrace prévoit d'intégrer les produits de Cado à sa plateforme de sécurité ActiveAI, afin d'améliorer la collecte de données dans les environnements cloud et d'optimiser l'analyse via Cyber AI, un outil dédié à l'investigation et à la priorisation des alertes.
L'opération, dont le montant n'a pas été divulgué, devrait être finalisée en février, sous réserve de l'approbation des autorités réglementaires compétentes. Le fondateur de Cado, James Campbell, ainsi que son équipe, seront intégrés à Darktrace.

Maki lève 26 M€ pour ses agents d'IA destinés aux RH

La startup Maki a bouclé un tour de table en série A de 26 M€. Une opération menée par Blossom Capital, avec la participation de DST Global et d'investisseurs historiques tels que Frst, GFC et Picus Capital. Maki développe des agents d'IA destinés aux services RH pour optimiser certains flux de travail, comme le processus de présélection et d'entretien, accélérer les délais de recrutement et réduire le turnover. Maki assure que, par ce biais, elle garantit à ses clients des gains en efficacité qu'elle chiffre à plusieurs millions d'euros, sans plus de précision.
Avec ces nouveaux fonds, Maki entend accélérer sa roadmap produit, afin de fournir à ses clients davantage de capacités de personnalisation sur leurs agents d'IA. L'entreprise compte également embaucher entre 50 et 60 collaborateurs, et s'étendre aux États-Unis en ouvrant des bureaux à New York. Les États-Unis représentent actuellement 30 % de son activité.

La startup d'IA d'Elon Musk lève encore 6 Md$

Huit mois après une série B de 6 milliards de dollars, la startup d'intelligence artificielle xAI a de nouveau levé cette somme pour sa série C. Parmi les investisseurs figurent Andreessen Horowitz (A16Z), Blackrock, Fidelity Management & Research Company, Kingdom Holdings, Lightspeed, MGX, Morgan Stanley, OIA, QIA, Sequoia Capital, Valor Equity Partners, Vy Capital, ainsi que des acteurs technologiques majeurs comme Nvidia et AMD. xAI est désormais valorisé à plus de 40 Md$.
Grâce à ces nouveaux fonds, la startup veut notamment accélérer le développement de son infrastructure. L'entreprise a également annoncé qu'elle recrute activement. Elle s'est dotée d'une infrastructure de calcul de pointe avec son supercalculateur Colossus. Mis en service en septembre 2024, il intègre 100 000 GPU Nvidia Hopper et est utilisé pour entraîner la troisième génération de modèles xAI, Grok-3, dont le déploiement a pris du retard. La capacité de

Colossus devrait doubler pour atteindre 200 000 GPU Nvidia Hopper, grâce à l'utilisation de la plateforme réseau Nvidia Spectrum-X Ethernet.

Blackstone investit 300 M$ dans DDN

Le fournisseur de solutions de protection des données pour les environnements HPC et IA vient de recevoir un investissement de 300 M$ de la part de Blackstone Tactical Opportunities. Avec cette arrivée d'argent frais, DDN se trouve valorisé à 5 Md$. Les fonds vont servir à soutenir la forte croissance que connaît l'entreprise surfant sur la vague de l'IA. DDN revendique des milliers de clients et supporte 500 000 GPU de Nvidia dans de multiples secteurs d'activités. xAI et Lambda sont des utilisateurs.

La startup d'IA et de cybersécurité BforeAI lève 9 M€

La startup montpelliéraine BforeAI, spécialisée dans la cybersécurité prédictive, poursuit son développement. Après un tour de table en série A de 15 millions d'euros en avril 2024, elle a a annoncé avoir levé 9 millions d'euros auprès des fonds américains Titanium Ventures et Syn Ventures. Fondée en 2020 par Luciano Allegro, Luigi Lenguito et Sebastian Cesario, l'entreprise compte actuellement plus de 90 salariés.

Elle développe une technologie de détection et de blocage des cybermenaces destinée au secteur financier. BforeAI utilise l'analyse prédictive pour détecter les comportements inhabituels dans un environnement clients, ainsi que les domaines usurpés. Une fois qu'une menace est détectée, son outil PreCrime vient les bloquer. Il est en mesure de stopper les usurpations d'identité, les tentatives d'hameçonnage entre autres. Avec ces nouveaux fonds, la société prévoit de soutenir sa R&D et renforcer son offre.

L'ESET étend ses liens avec le fournisseur de solutions Open XDR

La nouvelle intégration renforcée avec la technologie Open XDR de Stellar Cyber concerne en particulier deux solutions d'Eset.

D'une part, Eset Cloud Office Security (ECOS) qui renforce la protection des environnements Microsoft 365 et Google Workspace contre les attaques. Et d'autre part, Eset Threat Intelligence (ETI) qui enrichit la détection des menaces grâce à 15 flux TAXII, comprenant notamment des données sur les rançongiciels et les menaces APT. Ces informations, issues du réseau mondial de télémétrie d'Eset, permettent aux équipes de sécurité de réagir plus efficacement aux incidents.

La combinaison des deux offres apportent plusieurs bénéfices aux équipes de sécurité. De multiples solutions Eset sont intégrées dans la plateforme Stellar Cyber. Ceci permet une surveillance globale de l'entreprise, couvrant terminaux, emails, cloud et réseaux. La collaboration entre les deux entreprises rationalise la gestion des menaces, alliant détection et résolution simplifiées. La plateforme Stellar Cyber exploite les données de surveillance mondiale d'Eset pour améliorer la détection et la réponse aux incidents.

IBM et SAP collaborent pour une nouvelle solution

Les deux sociétés ont rendu publique la future solution Rise avec SAP sur IBM Power Virtual Server, conçue pour transférer les workloads SAP S/4HANA d'IBM Power Systems sur site vers le Cloud dans un délai de 90 jours.

Rise avec SAP est une transformation guidée qui propose des services ERP Cloud orientés sur les résultats, et des plateformes pour accompagner les entreprises à repenser leur modèle opérationnel. Cette nouvelle offre SAP Hyperscaler, Rise avec SAP sur IBM Power Virtual Server, aide à réduire les risques et à améliorer le temps de migration de SAP S/4HANA d'IBM Power Systems sur site vers le Cloud en moins de 90 jours. Les deux entreprises, accompagnées d'un écosystème de partenaires, travaillent ensemble à aider les clients d'IBM à se transformer grâce à Rise avec SAP sur IBM Power Virtual Server, avec des solutions et des capacités combinées, ainsi que des efforts conjoints de commercialisation. Les clients de Rise avec SAP sur IBM Power Virtual Server, les partenaires commerciaux d'IBM, auront également accès à un ensemble d'outils appelé IBM Transformation Suite for SAP Applications, afin de les aider à démarrer la migration vers SAP S/4HANA et Rise avec SAP. Cet ensemble comprendra des logiciels et des services de pointe pour l'évaluation des environnements existants, la migration des données et du code, ainsi que l'automatisation des tests.

Rise avec SAP sur IBM Power Virtual Server sera disponible dans un premier temps pour les clients américains, au deuxième trimestre 2025. La disponibilité sera ensuite étendue tout au long de l'année 2025.

Hammerspace et Cloudian signent un partenariat

Les deux entités s'accordent pour proposer une offre pour la gestion des données non structurées à l'échelle.

La solution va combiner le système de fichier parallèle d'Hammerspace et le stockage objet HyperStore de Cloudian. Avec l'intégration des fonctionnalités d'orchestration des données d'Hammerspace et la plateforme objet sécurisée de Cloudian, les entreprises vont pouvoir unifier leurs données sur l'ensemble des environnements dans un seul espace de nommage, simplifiant ainsi la gestion et l'accès aux données non structurées, tout en bénéficiant des performances que nécessitent les environnements d'IA et HPC. Les entreprises vont ainsi profiter des fonctions de mouvements des données d'Hammerspace sur des protocoles standards comme NFS, SMB ou S3 vers la plateforme HyperStore de Cloudian qui

fournit une compatibilité à l'API S3 de haut niveau et des fonctions de sécurité et de protection des données sur de très larges volumes. La solution est immédiatement disponible sur les réseaux des partenaires des deux éditeurs.

Canonical lance un programme de partenariat sur les vulnérabilités open source

L'éditeur annonce Ubuntu Security Research Alliance Program, un programme de partenariat gratuit entre Canonical et les organisations spécialisées dans le scan de vulnérabilités open source.

L'objectif de ce programme est de garantir une plus grande transparence et standardisation des données sur les vulnérabilités, tout en renforçant la sécurité des utilisateurs d'Ubuntu grâce à une détection proactive des menaces. Les organisations qui opèrent ou développent des solutions de scan de sécurité sont invitées à manifester leur intérêt pour rejoindre ce programme. L'Ubuntu Security Research Alliance Program a été conçu spécifiquement pour les fournisseurs de solutions de recherche et de détection, afin d'améliorer la précision et l'utilité des informations sur les vulnérabilités. Le but est d'intégrer des conseils de remédiation directement dans leurs résultats pour les utilisateurs d'Ubuntu.

Le programme profitera aux clients communs d'Ubuntu et des solutions de scan de sécurité. Il donnera à ces derniers un accès simplifié à des informations précises sur les vulnérabilités et les correctifs disponibles pour tous les paquets Ubuntu, ainsi qu'à des rapports exacts sur tous les produits Canonical dans les résultats des scanners. Cette collaboration permettra également aux opérateurs de scanners de sécurité de réduire les faux positifs et de fournir des recommandations plus exploitables pour les étapes de remédiation des CVE. Les membres du programme auront également un accès anticipé aux futures feuilles de route d'Ubuntu pour se préparer aux changements d'outils et de processus accompagnant les prochaines versions.

Cohesity étoffe sa solution CERT

Cohesity Cyber Event Response Team (CERT) étend les services proposés avec des partenariats sur la réponse à incidents (RI).

En s'associant à des fournisseurs de solutions de RI de premier plan, tels que Palo Alto Networks Unit 42, Arctic Wolf, Sophos, Fenix24 et Semperis, Cohesity CERT complète le processus de RI traditionnel en y intégrant des données et une expertise en matière de sauvegarde et de restauration, ce qui permet d'accélérer les enquêtes et d'aider les clients à se remettre plus rapidement des incidents. La solution peut partager avec ses partenaires RI un ensemble consolidé de données opérationnelles approuvées par le client, notamment des journaux, des rapports, des inventaires, etc. Ce riche ensemble de données, associé à l'expertise approfondie de Cohesity CERT en matière de sécurité et de restauration des données, améliore les capacités de confinement des partenaires RI, d'investigation numérique (forensics), et de renseignement sur les menaces (threat intelligence). Ces données leur permettent d'effectuer une analyse plus efficace du cyber incident et de résoudre rapidement les problèmes tout en réduisant les temps d'arrêt d'activité. Les clients ont également la certitude que le partenaire RI de leur choix peut collaborer directement avec Cohesity pour rationaliser leur cyber-réponse et s'assurer qu'ils restaurent des données propres plus rapidement.

Valeo et AWS collaborent sur le SDV

Au CES de Las Vegas qui vient de débuter, Valeo a annoncé sa collaboration avec AWS autour du véhicule défini par logiciel.

La collaboration entre Valeo et AWS permet un développement, des tests et une validation plus rapides et plus efficaces des logiciels dans tous les domaines de l'automobile, tels que les systèmes avancés d'aide à la conduite (ADAS), les systèmes d'info-divertissement et la mobilité autonome. Elle contribue également au développement de nouvelles fonctionnalités pour l'utilisateur final pour une meilleure expérience de conduite. Les trois premières solutions présentées par Valeo au CES 2025 dans le cadre de cette collaboration sont le Valeo Virtualized Hardware Lab, le Valeo Cloud Hardware Lab et le système d'assistance Assist XR. Afin de renforcer la collaboration entre les deux entreprises, Valeo adhère au réseau de partenaires (APN) d'Amazon Web Services (AWS). L'APN est une communauté mondiale de partenaires AWS qui s'appuie sur les technologies, les programmes, l'expertise et les outils AWS pour créer des solutions et des services pour leurs clients.

AGENDA

Université des DPO
6-7 février 2025
Maison de la Chimie, Paris

WAICF
13-15 février 2025
Palais des Festivals, Cannes

Web Summit Qatar
23-26 février 2025
Doha Exhibition and Convention Center — Doha, Qatar

MWC
3-6 mars 2025
Fira Gran Via Barcelone, Espagne

Nvidia GTC
17-20 mars 2025
Convention Center San Jose, USA

Adobe Summit
18-20 mars 2025
The Venetian Convention and Expo Center — Las Vegas, USA

Qualtrics X4
18-20 mars 2025
Salt Palace Convention Center Salt Lake City, USA

Face à l'urgence, pour être prêts à tout, tout de suite, tout le temps.

Faites un don à notre Fonds d'urgence.

ALLIANCEURGENCES.ORG

1 CLIC, 1 DON, 6 ONG EN ACTION

ALLIANCE URGENCES · ACTION CONTRE LA FAIM · care · handicap international · MÉDECINS DU MONDE · PLAN INTERNATIONAL Jusqu'à l'égalité · solidarités international

Conformité
Quelles priorités pour 2025 ?

NIS 2, DORA, CSRD, facturation électronique… Les règlements et lois de conformité s'enchaînent et se bousculent dans l'agenda des entreprises. Sur des plans différents, elles demandent toutes l'attention des DSI et des instances dirigeantes. Mais entre toutes celles-ci, comment prioriser ce qui doit être fait maintenant et ce qui peut encore un peu attendre ? Avec quels outils démontrer cette conformité à la norme demandée ? Comment font les autres ?

C'est à ces questions que ce dossier veut répondre et aider les entreprises à mettre en place des véritables stratégies vis-à-vis de la cascade de règlements annoncés et ceux à venir.

Dossier réalisé par Bertrand Garé

Des **priorités** multiples

2024 a vu fleurir les attaques cyber, mais aussi les demandes pour que la résilience des entreprises s'améliore avec de nombreuses règles autour des données et des infrastructures des entreprises. Il est possible d'y ajouter un versant durable qui lui aussi concourt à un meilleur usage de cet ensemble. Mais aujourd'hui, comment les entreprises font-elles le choix entre toutes les réglementations en place ou à venir ?

Selon Maxime Maintier, directeur RSE chez Fujitsu : *« Alors que la directive NIS2 devait initialement secouer les stratégies de cybersécurité des entreprises, le report de sa mise en œuvre a engendré une réorientation temporaire des priorités. Cependant, cette accalmie ne doit pas faire baisser la garde. À l'approche de la date définitive, les entreprises devront s'atteler à renforcer leurs dispositifs pour se conformer aux exigences européennes, marquant une montée en puissance des projets de sécurisation. »*

Dans le secteur financier, Imad Abounasr, associé EY, expert cybersécurité sur le secteur des services financiers indique : *« Pour les entreprises et la mise en conformité DORA, l'une des priorités est réellement la mise en place de la gouvernance, au bon niveau de l'organisation, pour gérer l'ensemble des aspects liés aux risques TIC de façon holistique* (risque TIC, risque des tiers TIC). » Il ajoute : *« À l'inverse de RGPD, qui était vraiment nouveau à l'époque, DORA constitue une harmonisation de sujets déjà existants, il n'y aura donc pas de période de grâce et les premiers contrôles se feront dès le 17 janvier. Elle reste malgré tout un changement de paradigme qui vient intégrer la résilience numérique au cœur des activités des entreprises. »*

Antony Derbes, président d'Open Lake Technology, résume : *« À l'horizon 2025, les institutions financières devront faire face à un cadre réglementaire de plus en plus complexe et contraignant. Les nouvelles exigences imposées par les régulateurs, notamment en matière de résilience opérationnelle, d'intégrité des données et de cybersécurité, changent radicalement la manière dont les entreprises gèrent la conformité. Dans ce contexte, la mise en place de solutions technologiques adaptées devient cruciale pour garantir non seulement la conformité, mais aussi la compétitivité sur le long terme. »*

Des entreprises (trop ?) confiantes

Dans une étude réalisée pour le compte de Zscaler en juin dernier, les responsables informatiques semblaient convaincus que leur entreprise sera en mesure de respecter la directive NIS 2 avant la date butoir. Les quatre cinquièmes (80 %) des personnes interrogées sont de cet avis, tandis que 14 % des responsables interrogés affirment avoir déjà rempli leurs obligations il y a déjà plusieurs mois, donc avant la date limite.

Imad Abounasr,
associé EY

« Pour les entreprises et la mise en conformité DORA, l'une des priorités est réellement la mise en place de la gouvernance, au bon niveau de l'organisation »

Les équipes IT ne sont pas les seules à accorder de l'importance aux réglementations. Elles ont le soutien des équipes dirigeantes, conscientes du rôle majeur de ces réglementations dans le succès de leur cybersécurité. L'enquête démontre pourtant que cette confiance repose sur des bases fragiles. Seule la moitié des personnes interrogées (53 %) estime que leurs équipes maîtrisent parfaitement les exigences liées à la conformité à NIS 2. Un chiffre qui n'atteint plus que 49 % quand on leur demande s'ils ont le sentiment que les dirigeants ont réellement compris ces exigences. Le rapport a également révélé un décalage entre la présentation de la directive et la compréhension qu'en ont les responsables informatiques. NIS 2 se présente comme une directive destinée à renforcer la sécurité globale, et comme une extension du cadre NIS en vigueur. Mais près des deux tiers des personnes interrogées (62 %) estiment qu'elle représente une nette inflexion par rapport à leur stratégie actuelle. Ces chiffres indiquent que de nombreuses entreprises n'ont pas su anticiper l'évolution des solutions technologiques et se sont contentées de maintenir le strict minimum en matière de sécurité, aussi longtemps qu'elles l'ont pu. Les responsables informatiques ont identifié trois domaines spécifiques nécessitant des changements majeurs pour se conformer aux nouvelles exigences : l'actualisation de leur pile technologique ou de

leurs solutions de cybersécurité et la sensibilisation des collaborateurs et de la Direction. La directive comporte également trois aspects posant véritablement problème aux personnes interrogées : la sécurité des réseaux et des systèmes d'information (31 %), les règles de base de la cyber hygiène et la sensibilisation (30 %), ainsi que les politiques et procédures relatives à l'efficacité des mesures de gestion des risques en matière de cybersécurité (29 %). Il faut remarquer que ce sondage a été réalisé auprès d'entreprises de plus de 500 salariés, ce qui induit un biais important sur la vision globale que l'on peut avoir sur l'état de préparation des entreprises sur cette nouvelle norme, dont l'ensemble des règles n'a pas encore été clairement défini. Les retards dans la transposition des principales nouvelles règles de conformité sur NIS 2, DORA et CSRD amènent de plus une période d'incertitude qui ne mène pas à l'action immédiate. CRSD est en outre appelé à évoluer avec un nouveau « bus » de mesures visant à alléger le reporting demandé. De plus, l'annonce par l'ANSSI d'une période de transition de trois ans fait que l'avancement vers la norme est très disparate et varie selon la taille, la criticité de l'environnement et la volonté des dirigeants de l'entreprise.

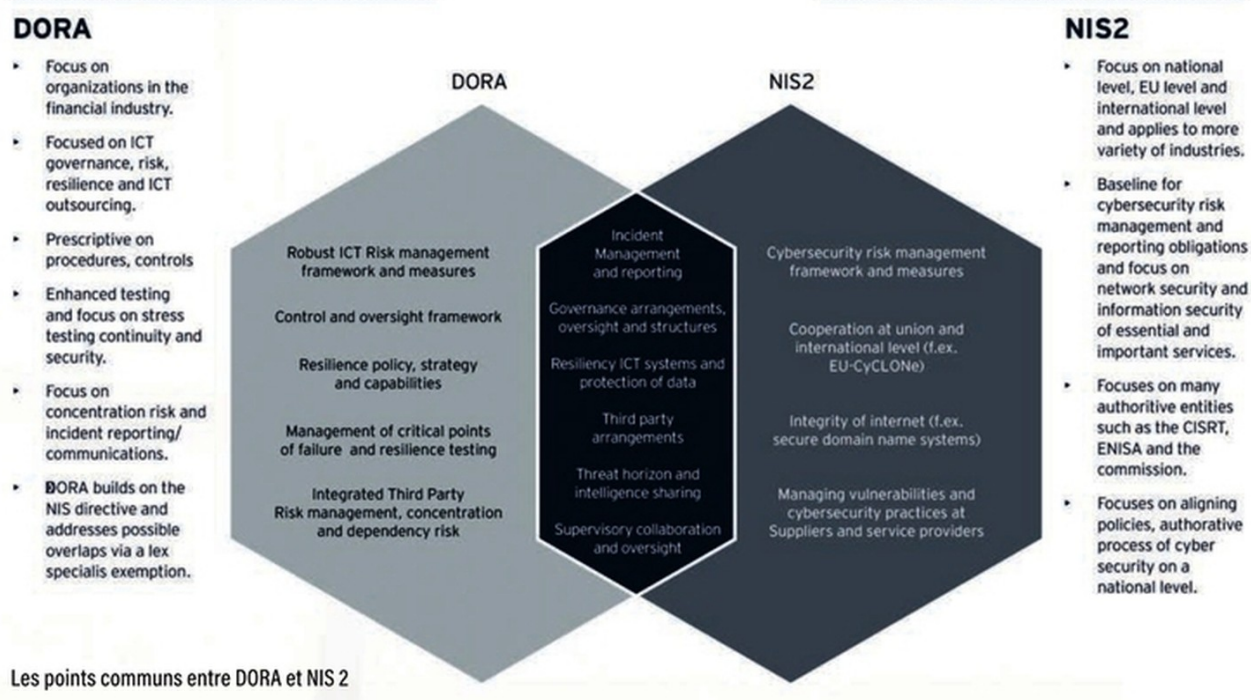
Des efforts importants à consentir

Benjamin Gras, avocat associé dans le cabinet Inside Avocats à Lille, résume assez simplement ce qu'il constate dans les entreprises qui l'interroge sur ces questions de conformité : *« du point de vue des régions, DORA est un sujet très... parisien. Cela concerne majoritairement les banques et le secteur de l'assurance, dont les sièges sont effectivement à Paris, ou des avocats du barreau de Paris qui seront principalement consultés, avant que cela soit décliné en région, dans les filiales »*. Il ajoute : *« la priorisation va être faite en fonction des enjeux financiers et aussi, pour ce qui concerne la partie données personnelles, en fonction des priorités de la CNIL qui sont annuellement données. La CNIL va donner ses secteurs d'activité pour 2026, même si elle a aussi annoncé son plan stratégique, qui comprend notamment l'IA et la cybersécurité, le RGPD et NIS 2 vont être des éléments très prioritaires. Je pense qu'on entendra vraiment parler de NIS 2 dans quelques années. Pourquoi ? Parce qu'il va y avoir un effet cascade. Il va falloir que les secteurs essentiels et les secteurs critiques se mettent en conformité. On rentre en fait toujours dans la même phase : 1/ j'audite, 2/ je déploie et je vérifie. Sur les secteurs essentiels, la vérification devra aussi être effectuée chez les sous-traitants, d'où cet effet de cascade qui va prendre du temps »*.

Imad Abounasr renchérit : *« il va y avoir encore un peu de run à faire sur la gestion du stock, avec notamment des enjeux autour des tiers, l'évaluation des tiers et la remédiation contractuelle. Vu le volume que cela représente, je ne connais aucune banque qui sera prête au 17 janvier 2025. Ils vont être prêts sur le socle, sur la base, sur les procédures, mais pas sur la remédiation complète. C'est la deuxième partie d'un chemin vers la conformité DORA »*. Ce point sur l'audit et la chaîne des acteurs autour de la résilience de l'entreprise est revenu dans quasiment tous les entretiens que nous avons eus lors des interviews pour réaliser cet article.

Les entreprises en retard sur le reporting CSRD

Alors qu'il est vu comme trop lourd et trop complexe, le reporting CSRD va certainement évoluer pour le rendre plus accessible. En l'état, une étude de Semarchy constate un



Les points communs entre DORA et NIS 2

retard patent des entreprises. Alors que 65 % des organisations éligibles aspirent à être prêtes pour l'audit et à respecter les échéances de reporting imminentes dans les 12 prochains mois, un quart (25 %) manque aujourd'hui de confiance dans la qualité et la fiabilité de leurs données ESG, et moins d'un tiers (27 %) pense disposer actuellement de la gestion des données et des systèmes nécessaires pour répondre aux exigences strictes en matière de reporting. Une organisation sur trois (31 %) déclare que l'incertitude persistante autour des lignes directrices l'incite à adopter une approche mesurée en matière de conformité.

Cette réponse s'inscrit dans un contexte de rumeurs selon lesquelles l'UE annoncera un ensemble simplifié de directives sur le développement durable en février 2025. 68 % des entreprises déclarent que les cadres supérieurs chargés des données, y compris les chief digital officers (CDO) et les directeurs de l'information (CIO), jouent un rôle central. Cela suggère un nouveau niveau de responsabilité pour les équipes informatiques, qui regroupent de vastes quantités de données provenant de sources disparates dans des rapports obligatoires utilisés par les investisseurs, les analystes et autres parties prenantes pour évaluer les performances et les risques d'une entreprise en matière de développement durable.

En comparaison, seulement 54 % des cadres affectés au développement durable ou à l'ESG sont concernés. Et seulement 33 % des entreprises interrogées déclarent que leur directeur financier est impliqué dans le processus de conformité malgré les lourdes implications financières des changements réglementaires. Plus des deux tiers (68 %) des entreprises prévoient d'allouer plus de 10 % de leur budget informatique annuel à la conformité CSRD, et plus d'un quart (26 %) prévoient même d'investir plus de 20 %. Il est à noter que 89 % des entreprises collectent et communiquent des données ESG depuis au moins un an, et plus de la moitié d'entre elles (58 %) le font depuis plus de trois ans. □

NIS 2 CHEZ LES EUROPÉENS

Belgique : Le projet de loi a été voté en avril 2024 et les modalités pratiques de mise en œuvre sont en cours de finalisation. Le Centre pour la Cybersécurité Belgique (CCB) a mis en place un cadre de référence basé sur la norme ISO 27001.

Allemagne : Le pays a publié plusieurs versions de son projet de loi, avec une transposition complète prévue pour 2025. Le cadre réglementaire s'appuie sur la loi KRITIS et le BSI Act.

Croatie : Le Croatian Cyber Security Act (CSA) est entré en vigueur en février 2024, permettant d'ores et déjà une transposition partielle de la directive.

Hongrie : La directive NIS 2 a été transposée en mai 2023, avec une législation hongroise exigeant l'enregistrement des entités essentielles et importantes avant juin 2024.

Les offreurs à la chasse aux opportunités

La cascade de nouvelles réglementations autour de la résilience fait le miel des offreurs du marché autour de la protection ou de la gestion des données. Que ce soit pour répondre à DORA, NIS 2 ou à CSRD, les éditeurs et ESN fourbissent leurs arguments pour convaincre les clients.

Nous en avions déjà parlé dans un numéro précédent. Datacore a étendu son offre à la cybersécurité pour répondre aux besoins de NIS 2. Dave Zubrowski, le CEO de Datacore expliquait : *« La directive NIS 2 ne concerne pas uniquement la conformité : elle vise également à renforcer la confiance dans l'écosystème numérique.Notre nouvelle solution de cybersécurité est conçue pour doter les entreprises des moyens d'anticiper, de résister et de se remettre des cybermenaces sophistiquées, tout en se conformant aux normes réglementaires. Nous nous engageons à aider nos clients à gérer leur entreprise en toute sécurité et en toute confiance, même dans un contexte de cyber-défis en constante évolution. »*

Pour sa part, TVH Consulting propose une offre dédiée à cette norme, NIS 2 Secure, qui propose un accompagnement GRC/AMOA pour aider les entités concernées à se mettre en conformité avec NIS 2, ainsi qu'une offre logicielle et de services complets. La solution comprend une analyse du contexte, basé sur des entretiens et une revue documentaire de l'existant. Suit la construction d'une feuille

de route pour atteindre la conformité sur trois ans sur tous les SI. Un accompagnement en mode RSSI externalisé, en gestion de projet ou au forfait, complète le dispositif avec des livrables et des tâches bien définies. Elle comprend, de plus, la mise en place d'une solution logicielle SMSI conforme aux exigences de NIS 2, avec APOS, outil de pilotage du SMSI et de la gouvernance de la cybersécurité à destination des RSSI. Disponible en mode SaaS ou sur site, la solution permet une évaluation continue des risques pesant sur le système d'information et les données de l'entreprise. Elle facilite le pilotage des incidents de sécurité et celui des audits de conformité, comme ceux requis par la norme ISO27001, ainsi que des programmes de formations certifiantes.

L'offre « NIS 2 Ready », lancée par SysDream, a vocation à accompagner les RSSI de bout en bout, en leur fournissant une feuille de route sur les différentes étapes à suivre pour accompagner la mise en conformité de leur entreprise sur une période d'un à deux ans. Cette feuille de route inclut notamment la réalisation de certains types d'audits, la formation et la sensibilisation des collaborateurs, ainsi que la mise en place de solutions adaptées à l'activité de l'entreprise et à son architecture IT en matière de détection et de réponse aux incidents.

Les coûts financiers de chaque étape seront également évalués séparément, ce qui permettra aux RSSI de décomposer la démarche, et ainsi de la rendre plus intelligible aux yeux de leur direction, notamment en termes de risques, de temps requis pour une mise en conformité optimale et d'investissements à planifier. Cette offre sera bien entendu modulable en fonction de l'activité de l'entreprise, son caractère stratégique et son exposition à d'éventuelles cyberattaques. Enfin, cette feuille de route permettra également aux RSSI de mieux comprendre leurs obligations et la démarche à suivre pour être conforme à cette nouvelle directive, et de repenser efficacement leur stratégie cyber en planifiant leurs travaux sur le long terme.

DORA pour aller plus loin que l'exploration

Dynatrace lance l'application Compliance Assistant, spécialement conçue pour fournir aux organisations la visibilité, les informations et l'automatisation nécessaires pour atténuer les risques et réduire les vérifications fastidieuses de configuration de la conformité associées à la loi DORA.

Azul, la seule entreprise entièrement dédiée à Java, indique que les pratiques de gestion des risques intégrées à ses solutions OpenJDK répondent pleinement aux exigences de stabilité, de résilience et d'intégrité stipulées par le règlement européen sur la résilience opérationnelle numérique (DORA). À l'approche rapide de l'échéance d'application, des milliers d'organisations financières européennes et d'entreprises opérant dans l'UE doivent agir pour garantir que leur infrastructure informatique respecte des normes rigoureuses en matière de résilience opérationnelle, ce qui peut nécessiter des investissements de temps considérables. Les versions Java de support à long terme (LTS) d'Azul visent à garantir une stabilité et des mises à jour de sécurité continues — y compris pour les anciennes versions comme Java 6 et 7 — essentielles pour maintenir la résilience opérationnelle sous la surveillance réglementaire. Les fonctionnalités de sécurité de l'entreprise, les tests approfondis et la compatibilité avec les architectures modernes et les environnements cloud offrent ainsi une plateforme Java sécurisée et évolutive.

Datadog, fournisseur de la plateforme de monitoring et de sécurité des applications cloud, annonce que sa solution de cadres règlementaires est configurée pour simplifier la réglementation Digital Operational Resilience Act (DORA), afin de fournir aux clients une visibilité immédiate et en temps réel des erreurs de configuration de leur environnement qui affectent leur audit DORA.

En tant que prestataire de services TIC, Claranet est directement concerné par DORA, et met en œuvre les mesures techniques et organisationnelles de sécurité adaptées aux besoins du client. L'entreprise propose une variété d'audits de sécurité, tels que l'analyse de la surface d'attaque, les scans de vulnérabilité, les tests de segmentation des réseaux, l'audit de la configuration cloud et les tests de pénétration (pentest). Ces audits visent à accélérer la cartographie des services TIC, et à éprouver l'efficacité des dispositifs de sécurité des entités financières.

NetApp, pour sa part, a mis en place un nouveau service dans la région EMEA : le Professional Services DORA Enablement Consulting Program. Celui-ci a été conçu pour aider les entités financières concernées par DORA et leurs prestataires de services à se conformer à la loi européenne DORA.

L'incertitude autour de CSRD

Si la norme et ces exigences risquent fort d'évoluer, les offreurs du marché IT sont d'ores et déjà prêts à y répondre. L'éditeur allemand SAP a ajouté Green Ledger, un système de comptabilité durable, à sa suite dédiée RSE. Le logiciel attribue les émissions de carbone à des activités économiques spécifiques, des transactions capturées par le système ERP et les applications financières de SAP. En intégrant, à ces solutions, les données sur les émissions, les entreprises peuvent prendre des décisions plus éclairées et durables, liant l'impact environnemental à la performance financière, tout en renforçant la conformité, l'efficacité et la transparence.

Une vue de SAP Green Ledger

SAP Green Ledger a été développé avec le soutien de partenaires tels qu'Accenture, Deloitte, EY et TCS (Tata Consulting Services), ainsi que des clients pilotes comme Covestro qui évalue actuellement SAP Green Ledger dans une phase pilote anticipée et teste la liaison des valeurs de dioxyde de carbone à SAP Green Ledger, car elles sont générées lors de la fabrication de produits spécifiques dans la chaîne d'approvisionnement.

Cegid a développé un connecteur CSRD/RSE compatible avec un ERP. Ce connecteur CSRD/RSE est déployé sous forme de module intégré à la solution Cegid XRP Ultimate. Il facilite la collecte de la data financière et extra-financière disponible au sein de l'ERP de Cegid. Avec l'intégration de ce module conforme aux enjeux de la directive européenne CSRD, les directions financières des ETI, grands comptes et entreprises d'intérêt public pourront automatiser la collecte d'une partie des données indispensables à la tenue du reporting extra-financier. □

Des **exemples récents** de mise en œuvre

Que ce soit pour de nouvelles normes ou celles déjà existantes, les entreprises se doivent de rester en conformité sur bien des points. Des exemples récents éclairent la manière dont elles s'y prennent pour rester dans les clous.

Le laboratoire Servier va s'appuyer sur les logiciels de Veeva, pour harmoniser ses opérations réglementaires à l'échelle internationale. En regroupant les processus réglementaires sur une plateforme unique, Servier pourra améliorer la procédure d'enregistrement de ses traitements en termes de rapidité et de conformité réglementaire. L'harmonisation des données, des documents et des processus réglementaires sur Veeva RIM permettra à Servier de rationaliser ses opérations et d'accélérer l'accès aux nouveaux médicaments pour les patients. Veeva RIM est une plateforme unifiée d'applications réglementaires, comprenant Veeva Registrations, Veeva Submissions, Veeva Submissions Publishing, et Veeva Submissions Archive, qui offre aux laboratoires pharmaceutiques comme Servier une visibilité en temps réel tout au long du processus réglementaire. La solution permet de s'adapter rapidement à l'évolution des exigences réglementaires de l'Union européenne. Veeva s'associe à Servier pour déployer la technologie dans toute l'Europe et dans le monde entier.

Respecter les obligations métiers

SEA Finance devait se conformer aux exigences de l'AMF, imposant la conservation des enregistrements des conversations relatives aux ordres et à la mise en place de contrats pendant au moins cinq ans. La solution précédente ne répondait pas de manière optimale à cette exigence. SEA Finance rencontrait plusieurs défis critiques avant l'intégration de la solution ASC Recording Insight. SEA Finance recherchait également une solution conforme aux obligations réglementaires, une gestion simplifiée des enregistrements, accessible à tous les utilisateurs, l'élimination des serveurs dédiés et la possibilité d'enregistrer les conversations via des téléphones portables. Avant l'intégration de la solution d'ASC Technologies, SEA Finance utilisait une infrastructure

complexe avec des serveurs internes pour l'enregistrement des conversations. Cela posait des problèmes de coordination entre les prestataires de téléphonie et d'informatique, entraînant des inefficacités. Avec ASC Technologies, SEA Finance a pu créer une architecture informatique flexible en utilisant une solution cloud sécurisée. L'accès se fait via une URL menant à une interface web intuitive, permettant une connexion facile et sécurisée pour la gestion des enregistrements. Le partenariat de SEA Finance avec adista a été crucial dans la transition réussie vers la solution ASC Recording Insight. Grâce à leur expertise et leur soutien stratégique, adista a accompagné SEA Finance tout au long du processus, offrant des conseils avisés et une assistance personnalisée. Leur implication a été essentielle pour garantir une transition fluide et efficace.

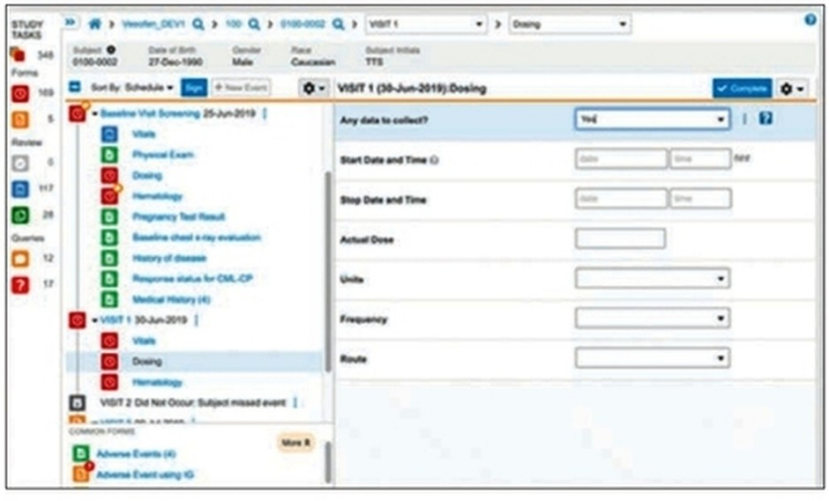
Un exemple dans le médical

La plateforme de gestion des effectifs de Lantum est utilisée par plus de 37 000 cliniciens pour soutenir la dotation en personnel de plus de 3 000 organismes de santé, y compris le National Health Service (NHS) du Royaume-Uni. En tant que principal utilisateur de Google Workspace, Lantum était préoccupé par le risque d'accès non autorisé aux informations stockées dans Google Drive et souhaitait être alerté de tout téléchargement de données sensibles. Avec le déploiement de Lookout Secure Cloud Access — la solution CASB (Cloud Access Security Broker) de Lookout — les équipes informatiques et de conformité de Lantum peuvent désormais surveiller l'utilisation de ses données à travers toutes les applications et les clouds, appliquer rapidement des politiques pour restreindre l'accès non autorisé et assurer la conformité. La mise en œuvre de Lookout Secure Cloud Access aide l'entreprise à assurer la conformité avec des réglementations telles que la norme ISO 27001 de l'Organisation internationale de normalisation et la norme Cyber Essentials du centre national de cybersécurité du Royaume-Uni, ce qui permet à Lantum de continuer à être un fournisseur du NHS, qui est le principal moteur de revenus de son entreprise.

Lookout Secure Cloud Access protège les données stockées dans toutes les applications cloud et SaaS. Qu'il s'agisse de partager les données en externe avec des partenaires ou en interne avec des employés, la solution offre au service informatique le contrôle et la visibilité nécessaires pour garantir que les données d'une organisation restent protégées à tout moment. Grâce à des politiques d'accès et de sécurité des données adaptatives combinées à des analyses avancées, Lookout permet aux entreprises de protéger leurs données contre les menaces internes intentionnelles, l'exfiltration accidentelle de données, les fuites de données à partir de comptes compromis et d'autres menaces avancées basées sur Internet, sans réduire la productivité des utilisateurs.

Manitou et le CSRD

Le spécialiste des équipements de manutention s'est associé avec MeltOne, un cabinet de conseil. Pour faire face aux défis de conformité réglementaire et de gestion des données financières liés à la CSRD (Corporate Sustainability Reporting Directive). La collaboration entre Manitou Group et MeltOne a débuté en 2018, à la suite de la refonte de l'analytique du groupe. Conscient de la nécessité de moderniser ses pratiques, le groupe a opté pour la solution CCH Tagetik pour optimiser ses processus d'élaboration budgétaire et de suivi du réalisé. Cette collaboration ouvre la voie à une série de nouveaux projets axés sur l'efficacité opérationnelle et la transparence financière. Manitou Group a élargi l'utilisation de la solution CCH Tagetik intégrée par MeltOne pour des aspects de gestion des ressources humaines, en introduisant des outils de suivi de la masse salariale et de pilotage des effectifs. Fort de ces succès, Manitou a décidé en 2020 de remplacer sa solution de consolidation statutaire par CCH Tagetik pour avoir une solution financière complètement intégrée. Cette transition, menée par MeltOne, a permis de rationaliser les processus financiers et d'améliorer la précision et la fiabilité des rapports financiers. En 2022, conscient des enjeux croissants en matière de responsabilité sociale des entreprises (ESG) et de reporting réglementaire (CSRD / Taxonomie), le groupe a lancé une réflexion approfondie pour choisir un outil capable de les accompagner pour répondre à ces nouveaux défis. Après cette première phase d'étude des solutions du marché, c'est le module ESG de l'outil CCH Tagetik qui a été retenu. Manitou Group a de nouveau travaillé avec les équipes MeltOne pour implémenter ce nouveau module. Les travaux autour de la data sont l'un des principaux défis lors de ce projet, notamment en vue de se conformer aux nouvelles réglementations (CSRD / taxonomie). □



Une vue de Veeva Vault

CES 2025

Des solutions compactes et évolutives pour l'IA et le HPC

L'édition 2025 du CES (du 7 au 10 janvier à Las Vegas) a mis en lumière des avancées importantes dans l'intelligence artificielle, les graphismes et le matériel. Nvidia, DeepMentor, ThunderSoft, Ugreen, Qualcomm, AMD et Gigabyte ont présenté des innovations dédiées à des secteurs clés comme la robotique, le calcul haute performance et le stockage intelligent.

Lors du salon, les exposants ont multiplié les nouveautés. Le géant Nvidia a marqué le CES avec des annonces ambitieuses, centrées sur l'IA et les graphismes. Jensen Huang, son PDG, a ouvert l'événement avec un keynote de plus de 90 minutes rassemblant plusieurs milliers de participants. *« Nous sommes entrés dans l'ère de l'IA physique »*, a-t-il affirmé. Cette nouvelle phase est incarnée par des innovations comme Cosmos, Project Digis et la série GeForce RTX 50. La plateforme Cosmos est dédiée à la robotique et aux véhicules autonomes. Avec des modèles génératifs capables de simuler des scénarios complexes, Cosmos permet aux développeurs de concevoir des systèmes capables de comprendre, de planifier et d'agir en temps réel. Cosmos, disponible en licence ouverte sur GitHub, est déjà adopté par des entreprises comme Uber et Xpeng. Toujours dans l'automobile, Nvidia a présenté la plateforme Drive Hyperion, construite autour du système AGX Thor. Elle combine IA générative et données synthétiques pour simuler des milliards de kilomètres de conduite virtuelle, contribuant à la sécurité et à l'efficacité des véhicules autonomes. Toyota, Mercedes-Benz et d'autres constructeurs ont déjà adopté cette solution.

© MC

Le CES (7-10 janvier, Las Vegas) a mis en avant des avancées importantes en matière d'IA, de robotique, de graphismes et de solutions matérielles compactes. Cette année, le salon a accueilli plus de 140 000 visiteurs et plus de 4 000 exposants.

L'ère des superordinateurs personnels compacts

Dans le domaine du graphisme, Nvidia a repoussé les limites avec la GeForce RTX 50 Series, portée par l'architecture Blackwell. La RTX 5090, dotée de 92 milliards de transistors, établit un nouveau standard en performances visuelles et IA, notamment grâce à DLSS 4. Cette technologie utilise l'IA pour générer plusieurs images par calcul, améliorant ainsi la fluidité et la qualité graphique des jeux. Jensen Huang a décrit cette carte comme *« une bête technologique, aussi élégante qu'efficace »*.

Nvidia a également dévoilé Project Digits, un superordinateur personnel compact basé sur le processeur GB10 Grace Blackwell. *« Chaque développeur, ingénieur ou artiste aura besoin d'un supercalculateur IA personnel »*, a expliqué le dirigeant. Digits, qui sera disponible en mai, intègre l'ensemble de la pile IA de Nvidia, promettant des capacités inédites pour les créateurs et chercheurs.

Intel et ses nouveaux processeurs

Intel a dévoilé ses nouveaux processeurs Core Ultra 200 Series. Michelle Johnston Holthaus, PDG par intérim de l'entreprise, a déclaré : *« Nous façonnons l'avenir de l'informatique personnelle avec plus de 400 fonctionnalités IA. »* Les processeurs Core Ultra intègrent des NPU (unités de traitement neuronal) qui accélèrent les applications IA tout en réduisant la consommation énergétique. Ces processeurs sont particulièrement adaptés aux charges de travail d'IA en périphérie, où ils surpassent leurs concurrents sur des métriques comme le traitement des médias et l'analyse vidéo.

Pour les entreprises, la plateforme vPro a vocation à s'imposer comme une solution incontournable. Elle offre des capacités avancées de gestion informatique et de sécurité renforcée, en partenariat avec Microsoft pour des fonctionnalités exclusives sur Windows 11. *« Les PC Copilot+ basés sur nos processeurs Core Ultra transforment la productivité et offrent une sécurité inégalée »*, a affirmé Pavan Davuluri, vice-président de Windows + Devices chez Microsoft.

Transformer la conception des puces par l'IA

DeepMentor, spécialiste taïwanais de la conception de puces IA, a démontré son rôle clé dans l'évolution de l'intelligence artificielle. Joe Fun, vice-président de l'entreprise, a expliqué comment DeepMentor aide les fabricants à transformer leurs idées en circuits intégrés. *« Avant de fabriquer une puce, il faut concevoir les algorithmes et prouver leur efficacité »*, a-t-il déclaré.

L'approche de DeepMentor repose sur la miniaturisation des designs et leur validation sur des plateformes comme FPGA avant de passer à la production à grande échelle avec des partenaires comme TSMC. *« Nous permettons aux entreprises de créer des puces optimisées pour des applications spécifiques, qu'il s'agisse de drones, de robots ou de systèmes de défense »*, a ajouté Joe Fun. Avec des solutions adaptables aux grandes entreprises comme Sony et Samsung, DeepMentor veut s'imposer comme un acteur incontournable dans le domaine des semi-conducteurs IA.

Développé par Ugreen, la plateforme NASync iDX est un système de stockage intelligent, doté d'un assistant IA intégré. Le fabricant promet *« de transformer la gestion des données pour les professionnels et les particuliers »*.

Compacité au service de l'IA et stockage

Spécialisé dans la vente de matériel aux entreprises, ThunderSoft a présenté deux produits : le Mini PC G1 Elite et le Mini PC G1 IoT. Le G1 Elite, basé sur un processeur Snapdragon X Elite, offre 45 Tera Operations per Second (Tops) de performance IA dans un format compact. Ses multiples ports, son efficacité énergétique et sa compatibilité avec Windows 11 et Linux en font une solution polyvalente pour les entreprises et les développeurs. Le G1 IoT, premier PC IoT sous Windows avec Snapdragon, cible les environnements industriels. Avec une garantie de support jusqu'en 2036, il promet une longévité exceptionnelle pour les projets utilisant des IoT, combinant puissance de calcul et gestion IA avancée.

De son côté, le chinois Ugreen a dévoilé le NASync iDX, une solution de stockage NAS intelligente intégrant un modèle de langage pour transformer le stockage en une expérience interactive. *« Ce n'est pas qu'un simple outil de stockage NAS ; c'est une solution qui comprend vos besoins »*, a expliqué Vivian Lu, responsable presse de l'entreprise. Le NASync peut organiser automatiquement des fichiers, supprimer les doublons et résumer des réunions grâce à l'IA. Conçu pour des usages personnels ou professionnels, il propose jusqu'à 160 To de stockage, idéal pour les équipes travaillant à distance.

Qualcomm, l'IA localisée pour plus de sécurité

Propulsés par le processeur Snapdragon X, les portables exposés par Qualcomm permettent d'exécuter des modèles IA localement, réduisant ainsi la dépendance au cloud. *« Cela améliore la sécurité des données tout en réduisant les coûts »*, a expliqué Quentin Ochoa-Luvaas, membre de l'équipe marketing. Avec une connectivité Wi-Fi 7 et une puissance IA de 45 Tops via Hexagon NPU, ces ordinateurs se destinent à des usages allant de la bureautique aux tâches d'IA avancées. Qualcomm a également mis en avant des ordinateurs portables ultralégers, illustrant sa volonté de démocratiser l'accès à l'IA. *« Nous sommes très enthousiastes à l'idée de lancer la puce Snapdragon X, qui est notre option la plus abordable à partir de 600 dollars. C'est un excellent prix pour les PC personnels, mais aussi pour les petites et moyennes entreprises »*, a souligné Quentin Ochoa-Luvaas.

Chez AMD, la nouveauté était le processeur MI325X, un accélérateur HPC offrant de hautes performances, notamment 2,6 pétaflops en FP8. Construit sur l'architecture CDNA de troisième génération, il est conçu pour les charges de travail exigeantes comme les simulations scientifiques et la formation de modèles IA massifs.

Une évolution pour les centres de données

Spécialisé dans les serveurs, Gigabyte a présenté les modèles G893-SD1-AAX5 et G893-ZD1-AAX5, conçus pour des performances IA. Basés sur la plateforme Nvidia HGX B200, ces serveurs intègrent huit GPU Blackwell et offrent une bande passante GPU de 1 800 Go/s. Avec le Gigapod, une solution de supercalculateur en rack, Gigabyte répond aux besoins des entreprises cherchant à former des modèles IA génératifs massifs. Ces solutions combinent puissance, évolutivité et efficacité énergétique, confirmant la place de Gigabyte dans la liste des leaders des infrastructures HPC.

Bref, ce CES 2025 n'était pas seulement focalisé sur des gadgets dont l'utilité est souvent douteuse. De Nvidia à DeepMentor, en passant par ThunderSoft, Qualcomm, AMD et Ugreen, les entreprises ont présenté des solutions professionnelles. Alors que l'IA et les performances extrêmes prennent leur essor, le futur de la technologie semble plus prometteur que jamais. Comme l'a conclu Jensen Huang : *« L'avenir ne fait que commencer. »* □

M.C

CES Preview

Dell relooke ses PC et les renomme

Dell Technologies annonce la simplification et le rebranding de son portefeuille de PC, avec de nouveaux modèles IA conçus pour l'informatique personnelle et professionnelle.

Dell Technologies simplifie son offre en proposant désormais trois grandes catégories de machines :
– Dell : conçu pour le jeu, l'école et le travail
– Dell Pro : conçu pour une productivité de niveau professionnel
– Dell Pro Max : conçu pour des performances maximales.

Dans chacune des catégories, au-delà de l'entrée de gamme, il existe un niveau Plus qui offre les performances les plus évolutives, et un niveau Premium qui offre le nec plus ultra en matière de mobilité et de design. Les nouveaux PC AI intègrent une technologie NPU pour offrir des performances d'IA adaptées aux exigences spécifiques de chaque charge de travail. Dell étoffe sa gamme Intel avec les processeurs Intel Core Ultra (série 2), étend son offre AMD avec les processeurs AMD Ryzen, et poursuit sa collaboration avec Qualcomm pour offrir aux utilisateurs le matériel répondant le mieux à leurs besoins. En combinant innovation matérielle et logicielle, Dell offre des possibilités étendues d'utilisation de l'IA Copilot+ sur PC, visant ainsi un plus grand nombre d'utilisateurs. Il n'a pas été précisé si la touche magique allait s'ouvrir à d'autres modèles que celui de Microsoft lors d'un entretien, même si Dell soutient plusieurs fournisseurs de modèles. Kevin Terwilliger, VP & GM Dell PC lines, ajoute : *« il y a encore de nombreuses possibilités autour de Copilot, et celui-ci va devenir un standard pour les PC. Il faut aussi remarquer les annonces qui ont été faites à Ignite avec la recherche sémantique à travers un travail d'inférence quotidien. Vous n'avez plus à rechercher un document ou un fichier par son nom. Les recherches se font par un nom dans le fichier, une image dans le fichier, une diapo spécifique dans une présentation. Sincèrement nous pensons que les personnes vont de plus en plus utiliser cette touche ».* Dell a aussi de nombreux partenariats avec d'autres éditeurs de logiciels, ce qui lui permet d'ajouter d'autres possibilités avec la touche IA.

Michael Dell sur scène, lors de cette preview des annonces du CES

La vue d'un Dell Pro 14 Premium

Une conception modulaire et durable

Le nouveau portefeuille de Dell privilégie une conception modulaire innovante et l'utilisation accrue de matériaux recyclés, à faibles émissions et renouvelables pour les produits et leurs packagings. Le nouveau port USB-C modulaire de Dell est une innovation technologique qui vise à remplacer la soudure par des vis sur le port le plus couramment utilisé. Les appareils Dell Pro et certains appareils Dell Pro Max sont les premiers PC portables professionnels du marché intégrant un port USB-C modulaire qui est jusqu'à quatre fois plus durable tout en facilitant les réparations. La carte mère et des cartes E/S bénéficient de la même conception modulaire pour faciliter l'entretien et contribuer à réduire les déchets électroniques. Elle garantit que les réparations effectuées sur ces ports couramment utilisés n'affecteront pas la carte mère. Une nouvelle technologie sur la chimie des batteries réduit de 80 % l'usage de cobalt.

Par ailleurs, Dell a aussi réalisé des annonces autour de sa gamme Alienware, plus orientée « Gamers », ainsi que sur de nouveaux écrans. □

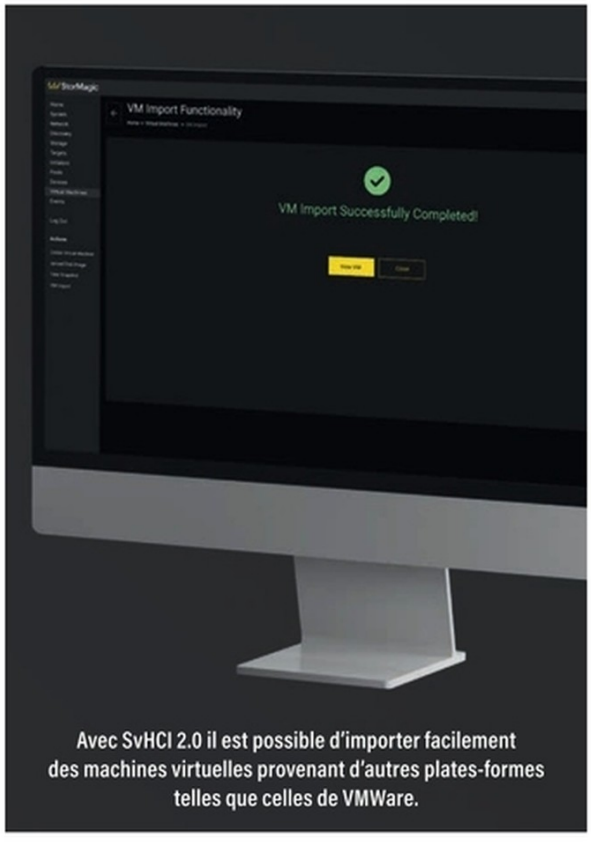
B.G

Edge Computing

StorMagic renforce sa solution hyperconvergente

Annoncée en juin dernier, la version 2 de SvHCI s'agrémente de nouvelles fonctionnalités.

Éditeur proposant des solutions de stockage pour les TPE-PME, StorMagic a pour mission de répondre aux besoins des entreprises en périphérie du réseau. Pour mieux correspondre à leurs besoins, StorMagic avait lancé une solution hyperconvergée combinant son logiciel SvSAN, une stack réseau virtuelle, son hyperviseur et un logiciel d'administration. La solution se voulait la moins chère du marché et visait clairement les entreprises utilisant VMware. La solution fonctionne sur deux serveurs de commodités, afin d'obtenir de la haute disponibilité. La sauvegarde reste aux mains de partenaires comme Veeam, Commvault ou Acronis. La solution peut se déployer sur site ou dans le Cloud avec la solution *« Witness as a service »*, un logiciel de contrôle à distance pouvant gérer jusqu'à 1 000 sites différents. Facturée au nœud, la solution est drastiquement moins chère que son équivalent chez VMware qui se rémunère au cœur de processeur. Pour la sécurité, les données sont chiffrées en transit et au repos selon la norme FIPS 140-2. La solution comprend son propre gestionnaire de clés ou s'intègre avec tout logiciel supportant KMIP.



Avec SvHCI 2.0 il est possible d'importer facilement des machines virtuelles provenant d'autres plates-formes telles que celles de VMWare.

Une résilience améliorée

Le premier axe d'amélioration de la V2 concerne la résilience de la solution, avec l'ajout du snapshot des machines virtuelles apportant une protection accrue des données par la possibilité de retour arrière après des mises à jour de l'OS ou de logiciels. La possibilité de restaurer les données dans un état antérieur après une attaque de rançongiciel. Il est possible de créer des snapshots à la main, en créant un point de restauration qui conserve les configurations du processeur, de la mémoire et du réseau. La solution peut supporter jusqu'à 16 snapshots par machine virtuelle. Par des mises en œuvre de politiques de rétention, il est possible d'automatiser les snapshots.

Une administration centralisée

Autre apport de cette nouvelle version, l'extériorisation de la console d'administration dans le cloud avec un contrôle sur l'ensemble des équipements hébergeant SvSAN ou SvHCI avec des rapports de fine granularité que ce soit par site ou par système. Outre l'hyperviseur de l'éditeur, la solution supporte vSphere, Hyper-V et KVM. Cette dernière fonction est le fondement de la troisième évolution d'importance de la plateforme avec l'import de machines virtuelles. Il est maintenant possible de migrer facilement et rapidement des machines virtuelles VMware vers les plates-formes de StorMagic. L'opération peut se réaliser directement depuis la console de SvHCI.

Par des « wizards », les migrations peuvent se réaliser par lot par des opérations simples. Après la connexion à l'instance ESXi, il suffit de sélectionner les machines virtuelles (un test automatique vérifie que cela est possible) et de choisir l'hôte SvHCI dans le cluster, puis d'assigner le réseau pour chaque machine virtuelle et de lancer l'opération de migration.

Autre fonction intéressante de la version, il est possible maintenant d'étendre le cluster sur deux localisations géographiques différentes par une extension en stretch cluster. De plus, le support des réseaux virtuels a été amélioré. La capacité des disques virtuels a été étendue avec le support de quatre disques virtuels par machine virtuelle, que ce soit sur des contrôleurs Virtio ou PV-SCSI. Le support du nombre de VM a lui aussi été amélioré, avec désormais 50 machines virtuelles en haute disponibilité par cluster. □

B.G

Quantique

Eviden affûte sa stratégie avec le finlandais IQM

Eviden héberge, dans son usine d'Angers, l'IQM Spark, un calculateur quantique à supraconducteurs. Une machine qu'il met à la disposition de ses clients dans le cadre de ses offres de conseils et de service, en complément de ses émulateurs quantiques Qaptiva.

Si la course au nombre de qubits entre IBM et Google fait régulièrement la une des médias, Eviden, ex-Atos, a fait le choix d'une stratégie radicalement différente. En 2016, Atos lançait la Quantum Learning Machine, un émulateur quantique destiné à tester des algorithmes quantiques sur des CPU classiques. Plusieurs dizaines de machines ont été livrées dans le monde et Eviden propose cet émulateur sous forme d'appliance, le Qaptiva 800, mais aussi en version logicielle pour supercalculateurs. Cette version permet de simuler de 41 à 100 qubits logiques sur une infrastructure HPC. L'émulateur est capable de simuler des QPU (Quantum Processing Unit) de technologies quantiques différentes et le niveau de bruit quantique de chacun. Cet émulateur permet notamment d'alimenter l'activité conseil et formation d'Eviden sur les sujets quantiques, mais, comme le souligne Cédric Bourrasset, directeur des activités HPC IA et informatique quantique d'Eviden : *« Notre objectif est d'aider les industriels à se préparer à l'arrivée du quantique. Ceux-ci sont confrontés à de nombreux défis parmi lesquels, au premier plan, la diversité des technologies de qubits en lice et les nombreuses startups présentes sur le marché. Les clients sont un peu perdus. Notre position est d'être des conseillers agnostiques vis-à-vis des infrastructures et des technologies, et d'aiguiller nos clients dans leurs choix. »*

IQM, le partenaire hardware d'Eviden dans le quantique

Le responsable Eviden ajoute qu'outre la simulation, les entreprises souhaitent pouvoir tester leurs algorithmes sur un « vrai » calculateur quantique. C'est précisément dans ce but qu'Eviden a installé dans son usine d'Angers un Spark, le calculateur quantique mis au point par le finlandais IQM. Cette startup, qui est passée un temps dans l'incubateur d'Atos, a déjà construit plus de 30 systèmes complets. Le système Spark déployé à Angers ne compte que 5 qubits, mais la startup a une roadmap très agressive, avec un modèle à 154 qubits prévu pour la fin de l'année 2025, le cap du millier de qubits à l'horizon 2027 et l'objectif du million de qubits d'ici 2033… IQM, qui a déjà levé 200 millions d'euros, s'apprête à annoncer une nouvelle levée de fonds pour tenir ce cap. Outre la mise au point de calculateurs quantiques de plus en plus puissants, la stratégie du Finlandais est de développer une couche logicielle pour attirer les développeurs, une approche qui rappelle celle qui a si bien réussi à Nvidia avec CUDA, lancé en 2007. *« Beaucoup d'acteurs du quantique ne travaillent que sur le volet hardware, mais nous travaillons aussi beaucoup sur le volet logiciel »*, explique Mikko Valimaki, co-CEO d'IQM, *« nous construisons notre propre stack logiciel, une plateforme pour que nos partenaires puissent développer des algorithmes »*. L'analogie avec le californien et les GPU ne s'arrête pas là, puisqu'il vise une intégration avec le HPC, à l'image des GPU qui ont réussi à se faire une place dans tous les supercalculateurs modernes. □

A.C

Cédric Bourrasset, directeur des activités HPC IA et informatique quantique d'Eviden

« Avec la mise en place de l'IQM Spark, nous voulons disposer de capacités de calcul quantique, mais mettre simplement à disposition des QPU n'a que peu d'intérêt pour nous. Il s'agit d'apporter à nos clients un accompagnement algorithmique et de mettre à disposition des capacités de calcul "on-demand" que nous allons valoriser, mais uniquement dans le cadre de projets »

Dev : 2025 sera-t-elle l'année de l'audace autour des projets d'IA ?

Par Laurent Doguin, responsable DevRel chez Couchbase.

Chaque jour, chaque secteur d'activité découvre des usages et des outils d'IA susceptibles d'avoir un impact majeur sur leur quotidien et productivité. Malgré cela, certains peinent à passer à la mise en pratique, à l'image des développeurs.

Si les chatbots dominent les applications d'IA, le potentiel de transformation s'étend bien au-delà de ces derniers. 2025 sera-t-elle donc l'année du passage à l'action pour les développeurs ? Voici quelques tendances expliquant leur positionnement et leurs priorités pour les mois à venir.

La prudence reste de mise...

Par nature, les développeurs règlent des problèmes, mais leurs employeurs restent frileux quant à l'expérimentation de l'IA générative afin de les régler, et la gestion de des failles potentielles qui en découleraient. La crainte d'une atteinte à la réputation notamment dans des domaines sensibles fait partie des facteurs majeurs de dissuasion. Ainsi, les entreprises adoptent une approche passive et observent les succès et les échecs des concurrents. Alors que la confiance est primordiale dans la relation client, elles confirment leur intention de ne pas vouloir d'une IA imparfaite suscitant du scepticisme et des doutes.

...mais le besoin d'une assistance IA contrôlée devient pressant

Lorsqu'ils conçoivent des outils d'IA pour les développeurs, les éditeurs doivent veiller avant tout à simplifier l'expérience et privilégier la simplicité. Ce public recherche en effet des solutions à faible coût, facile d'utilisation et d'intégration, sécurisées et bien évidemment fiables. Limiter les générations de données non fiables, les hallucinations, sont une priorité pour les développeurs. Traiter ces problèmes potentiels représente un coût important. Ainsi, l'avenir réside probablement dans l'utilisation de solutions clé en main comprenant des workflows de RAG, de catalogue d'agent prêt à l'emploi, et de modèles plus petits que nos LLMs actuels, plus restreints mais spécialisés, adaptés à des tâches spécifiques.(SLM)

La montée en compétences pour soutenir la compréhension de l'IA

Certains développeurs juniors s'appuient sur des outils d'IA comme Github copilot pour être plus productifs et contrôler leur travail ou même en déléguer une partie. Ils ont ainsi une approche totalement différente de leurs homologues seniors. Les développeurs expérimentés se concentrent sur le pourquoi une solution fonctionne, tandis que les moins aguerris peuvent ne pas s'intéresser et comprendre pleinement le raisonnement qui la sous-tend si celle-ci a été générée par une IA. Si l'assistance de l'IA améliore l'efficacité, elle peut ainsi également créer un déficit de connaissances. Les entreprises doivent donc investir dans des stratégies de formation qui comblent ce fossé, en veillant à ce que les développeurs juniors acquièrent les compétences pratiques qui les aideront à progresser dans leur carrière. Ironiquement, l'IA elle-même peut jouer un rôle dans cette formation, en offrant des expériences d'apprentissage immersives et en temps réel qui guident les développeurs tout au long des processus de codage.

L'avancée de l'entraînement de l'IA ouvre à la voie à de nouveaux possibles

Alors que l'IA générative est fortement plébiscitée, d'autres technologies qui se sont généralisées ces derniers mois pour la soutenir et la rendre plus efficace, conserveront un usage indépendant. C'est le cas de la recherche vectorielle, rendue populaire pour la génération augmentée de récupération (RAG), et qui s'avère être un puissant accélérateur pour les fonctionnalités de recherche multimodale, sémantique et hybride au sein des principales bases de données. Cela traduit l'effet d'entraînement de l'IA, où les innovations dans un domaine peuvent en influencer d'autres de manière inattendue.

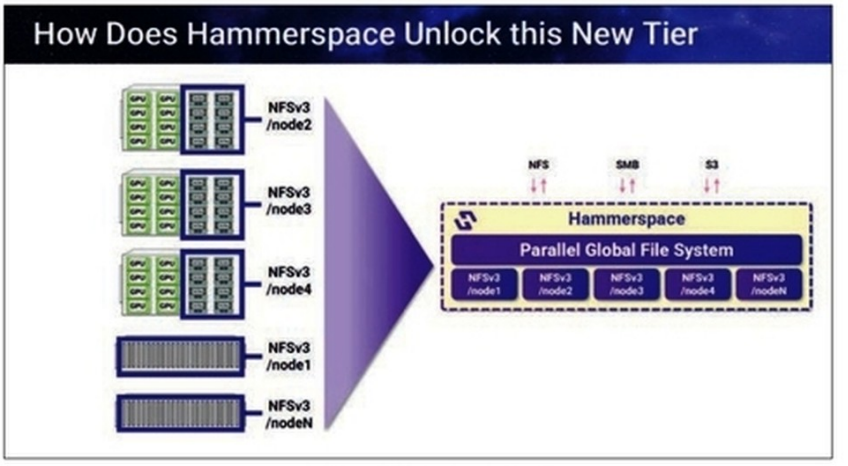
Si l'avenir de l'IA dans le développement peut sembler flou, les spécialistes prendront davantage confiance en leur capacité en 2025 et au-delà. Malgré une interaction complexe et une prudence de mise, la clé réside dans la mise en avant de la confiance, de la transparence et de l'éducation, en veillant à ce que l'IA accompagne l'autonomie des développeurs plutôt que de créer une dépendance qui entrave la compréhension de leur travail au quotidien et bride à terme l'innovation. Les choix que feront les développeurs ces prochains mois façonneront donc grandement l'avenir du secteur. Reste à voir lesquels seront-ils... □

Optimisation

Hammerspace optimise le traitement des charges IA

Lors du dernier Techlive d'A3 technology qui s'est tenu à Londres, Hammerspace a présenté la version 5.1 de son logiciel, ainsi qu'une amélioration technique créant un tier 0 pour un stockage partagé très rapide.

Le Tier 0 en étant activé, la plateforme de données d'Hammerspace débloque le stockage local en NVMe et délivre directement les données aux puces graphiques, chargées du traitement des charges d'intelligence artificielle à la vitesse du NVMe local. Les entreprises peuvent ainsi s'appuyer sur cette ressource partagée de haute performance, tout en conservant les gages de redondance et de protection des données que procure un stockage externe classique. Cette intégration fait du serveur local NVMe une partie du système de fichier global parallélisé et unifié se déployant sur tout type de stockage avec une orchestration des données automatique. La solution a, de plus, l'avantage de réduire significativement les coûts en rendant inutile d'avoir des stockages externes couteux.



La méthode déblocage du Tier 0

La solution vise à changer la donne dans la manière de configurer les architectures dans les traitements de données de l'intelligence artificielle, en rendant plus rapide et plus directe la possibilité d'utiliser pleinement les GPU à leur pleine puissance, tout en réduisant la consommation électrique et sans avoir à ajouter de ressources nouvelles.

Une kyrielle de nouveautés

La version 5.1 de la plateforme d'Hammerspace apporte son lot de nouvelles fonctions, avec le but d'améliorer la performance, la connectivité et la simplicité d'usage. Ainsi, la nouvelle version supporte nativement S3 pour le côté client en fournissant les permissions et l'espace de nommage par SMB et NFS. Les performances ont été largement améliorées. L'éditeur indique un facteur 2 pour les métadonnées et un facteur 5 sur la mobilité des données, l'assimilation des données en place et le cloud bursting. Des fonctions de résilience ont été ajoutées pour assurer la haute disponibilité du serveur de métadonnées. De nouvelles politiques d'automatisation étendent les « objectives » déjà présentes.

Les autres améliorations de la version concernent l'interface utilisateur avec des dashboards sous forme de tuiles personnalisables et des contrôles plus riches. Par ailleurs, le client SMB a été amélioré et des fonctions de haute disponibilité sont disponibles pour un déploiement dans Google Cloud. □

B.G

HAMMERSPACE ET CLOUDIAN SIGNENT UN PARTENARIAT

La solution va combiner le système de fichier parallèle d'Hammerspace et le stockage objet HyperStore de Cloudian. Avec l'intégration des fonctionnalités d'orchestration des données d'Hammerspace et la plateforme objet sécurisée de Cloudian, les entreprises vont pouvoir unifier leurs données sur l'ensemble des environnements dans un seul espace de nommage, simplifiant ainsi la gestion et l'accès aux données non structurées, tout en bénéficiant des performances que nécessitent les environnements d'IA et HPC. Les entreprises vont ainsi profiter des fonctions de mouvements des données d'Hammerspace sur des protocoles standards comme NFS, SMB ou S3 vers la plateforme HyperStore de Cloudian qui fournit une compatibilité à l'API S3 de haut niveau et des fonctions de sécurité et de protection des données sur de très larges volumes. La solution est immédiatement disponible par les réseaux des partenaires des deux éditeurs.

Infrastructure

Spie ICS veut marquer sa différence

Si dans l'inconscient collectif, le nom de Spie est associé à un secteur d'activité particulier, ce n'est plus la réalité. Séparé de Spie batignolles depuis des années, Spie ICS est dans un groupe spécialisé dans la fourniture de services multi-techniques et numériques.

Détenteur chez Spie de la pratique informatique, Spie ICS représente 15 à 20 % de l'activité du groupe Spie en France. Actuellement, le groupe développe une nouvelle stratégie sur trois ans. Xavier Daubignard, directeur général de Spie ICS, précise : *« Dans le cadre de ces plans stratégiques, on a coutume de déterminer quels seront nos moteurs de croissance pour les trois ans qui viennent, des thématiques sur lesquelles on est déjà mature et sur lesquelles on souhaite continuer à accélérer, et également des leviers de transformation qui sont des disciplines un peu plus émergentes sur lesquelles on souhaite investir pour continuer à transformer l'entreprise et à mieux répondre à nos clients. »* Il ajoute : *« Dans ce plan 2025-2027, les trois moteurs de croissance sur lesquels nous sommes déjà bien présents sur le marché sont l'infogérance globale d'une part, le cloud hybride d'autre part et la cybersécurité. Nous fixons aussi dans nos plans stratégiques, ce que l'on appelle des leviers de transformation. Les trois leviers de transformation sur lesquels nous souhaitons investir pour les trois années qui viennent, de manière beaucoup plus affirmée, sont l'automatisation des infrastructures, parce que Spie CS reste globalement positionné comme un acteur clé des infrastructures. Mais aussi l'IA qui, à l'origine, a été mise dans notre plan stratégique avec la vocation d'améliorer nos processus internes. Mais, en présentant nos initiatives internes à nos clients sur l'IA, on s'est rendu compte que cela pouvait les intéresser, qu'on pouvait leur apporter des solutions et nous avons décidé de développer une pratique IA à destination de clients en complément de l'usage de l'IA qu'on peut s'appliquer à nous-mêmes pour gagner en productivité. Le dernier thème sur lequel on se veut très actifs, c'est le thème de la décarbonation. On ne peut pas, en tant que Spie ICS, filiale de Spie qui est un acteur de la transition énergétique, et en tant qu'entreprise du numérique, regarder comment on peut décarboner le numérique ou comment on peut aider les autres filiales du groupe à décarboner les métiers de leurs clients par le levier numérique. »*

Xavier Daubignard,
directeur général
de Spie ICS

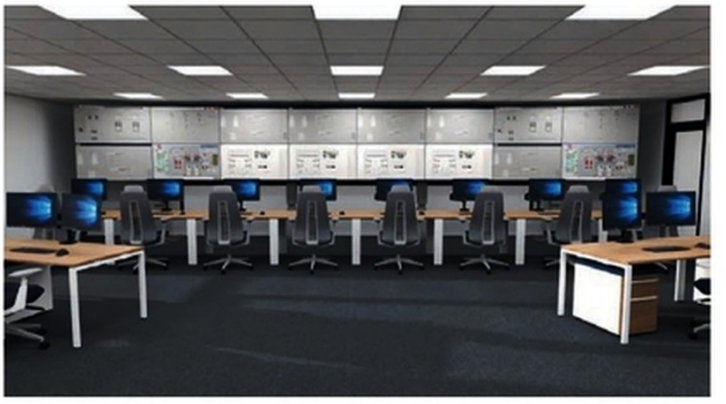
« Dans ce plan 2025-2027, les trois moteurs de croissance sur lesquels nous sommes déjà bien présents sur le marché sont l'infogérance globale d'une part, le cloud hybride d'autre part et la cybersécurité »

Les nouveaux bureaux d'ICS à Malakoff

Un acteur historique de l'infrastructure

L'ESN revendique sa maîtrise des infrastructures : *« on est un acteur historique des infrastructures, alors au sens étendu, on est présent sur le workplace, on est présent sur les infrastructures de réseau, de sécurité, cybersécurité, sur les infrastructures data center et, historiquement, dans une moindre mesure, sur le data management d'autre part. En étant un spécialiste des infrastructures, on a toujours très bien tiré notre épingle du jeu en termes de rentabilité. On a toujours suivi le mouvement du marché, et aujourd'hui, on voit que les infrastructures doivent être adressées en*

prenant en compte la « cloudification » *et le recours aux hyperscalers. On se rend compte qu'on doit effectivement adresser ce monde-là avec une forme d'hybridation des solutions qu'on pilote pour nos clients, c'est ce que l'on fait »*, explique Xavier Daubignard. Il précise : *« le tout cloud a été une mode, mais il y a un mouvement de balancier qui est en train de s'opérer, il y a un legacy qui est quand même extrêmement important chez les clients. Ce legacy combiné au cloud est une opportunité pour les acteurs comme nous. Bien savoir combiner ces environnements pour apporter un service sans couture aux clients, c'est quelque chose qui est extrêmement recherché aujourd'hui, et en étant un acteur français souverain, toujours à l'écoute des clients et qui apporte de l'expertise au plus près des besoins, on s'en sort plutôt très bien ».*



Une vue du centre de services managés de Spie ICS près de Grenoble

Une offre souveraine

« Plus de la moitié des collaborateurs de Spie ICS interviennent dans des secteurs contraints comme la défense ou l'aéronautique, qui sont nos secteurs principaux. On se doit, au titre de notre activité sur ces secteurs-là, de pousser la logique de souveraineté qui est, de mon point de vue, de plus en plus importante dans un monde qui est pour le moins chahuté. Après, il faut que la souveraineté puisse aussi s'exercer. Au regard du panel technologique que nous avons à notre disposition, qui vient souvent soit d'outre-Atlantique, soit d'Asie, on peut se poser des questions. Mais c'est notre parti pris en tant qu'entreprise française. L'ensemble de nos activités, de nos centres de services sur les sujets d'infrastructures sont positionnés en France, et nous nous sommes toujours refusés, pour faire vivre cette logique de souveraineté, à les externaliser, ne serait-ce qu'en nearshoring à l'étranger. C'est vraiment l'une de nos marques de fabrique. La limite de l'exercice, c'est effectivement qu'elle arrive au niveau du choix technologique. On est comme tout le monde, on utilise les plateformes technologiques qui sont à notre disposition, qui sont souvent des plateformes technologiques américaines. En revanche, dans la mise en œuvre, on fait en sorte de garantir une forme de souveraineté aux clients également », assure le DG de Spie ICS.

Sur la voie de la transformation

« Aujourd'hui, on se focalise sur trois nouveaux leviers de transformation. L'automatisation, j'en ai parlé, qui devient de plus en plus importante pour être performant dans le monde des infrastructures, en apportant des services d'automatisation à nos clients. À ce titre-là, nous venons de développer une plateforme agnostique et ouverte, pour traiter des cas d'usage réseau en automatisation. Il nous paraît essentiel de proposer une plateforme agnostique basée sur de l'open source, afin de pouvoir adresser l'automatisation en faisant fi de l'hétérogénéité technologique que l'on retrouve chez les clients. En parallèle, nous travaillons aussi sur l'automatisation d'un point de vue plus académique, puisque nous avons lancé une chaire de recherche sur l'automatisation des réseaux avec l'INSA à Lyon. C'est vraiment un sujet sur lequel nous voulons massifier nos investissements, compléter nos compétences pour pouvoir proposer à nos clients des infrastructures managées les plus automatisées possibles. Cela va s'étendre avec l'IA, et en particulier avec le projet d'automatisation de l'infrastructure. Les premiers cas d'usage que nous avons choisis sont liés au réseau, puisqu'on avait des cas clients qui nous ont été soumis, notamment par un très grand acteur de l'aéronautique européen, qu'on a dû adresser pour traiter ce sujet d'hétérogénéité. Nous nous sommes concentrés, dans un premier temps, sur les problématiques réseau. Si la promesse est tenue, cette plateforme d'automatisation a vocation à être étendue sur les sujets d'infrastructure data center, voire à intégrer des fonctionnalités autour de l'IA. Chez Spie, nous sommes pragmatiques, nous faisons donc de l'innovation pragmatique en validant sur la base. Si les promesses sont tenues, nous déploierons sur la plateforme. »

« L'autre thème, c'est l'IA. Nous sommes en train de décliner chez nos clients et chez nous notre stratégie IA pour tous, des agents génériques et des assistants personnels génériques d'IA. Pour délivrer ce service, nous avons marketé un produit qui s'appelle Orion. Nous l'avons lancé il y a 12 mois, à destination des 3 000 salariés de Spie ICS. Nous sommes très contents du succès de cette plateforme en interne chez ICS. Nous venons de prendre la décision d'étendre cette plateforme aux 19 000 salariés de Spie en France. Nous avons également marketé le produit, pour faire en sorte que l'on commercialise cet agent générique du quotidien auprès de nos clients. Nous avons quelques POC en cours chez certains de nos clients », indique Xavier Daubignard. Cela s'étend avec le développement d'agents spécialisés.

Dernier axe de transformation, la décarbonation, de l'IT, mais également du métier des clients par le levier IT. *« Nous avons été une entreprise plutôt précurseure dans la labellisation numérique responsable, puisque c'est, pour nous, une réalité depuis 2021. Nous avons été la première ESN du top 20 à être labellisée. Nous sommes désormais labellisés numérique responsable niveau 2, et nous souhaitons être une entreprise de référence sur le fait d'accompagner les clients vers un numérique plus responsable, vers des infrastructures plus responsables »*, met en exergue le dirigeant. ☐

B.G

Le Yoyo des cryptos

par Bertrand Garé

Le grand jeu de la spéculation autour des cryptos bat son plein. A la suite de son élection, Donald Trump avait indiqué vouloir mettre plusieurs cryptomonnaies dans une réserve. Celles-ci ont alors connu une hausse extraordinaire. Il a ensuite lancé sa propre cryptomonnaie, suivie par son épouse qui s'est prise au jeu, entraînant de nombreuses personnes à leur emboîter le pas en achetant ces cryptomonnaies au plus haut. Malgré tout cela, le Bitcoin, la plus emblématique de ces presque monnaies, avait du mal à battre son record de valeur. Il a suffi que le même Donald Trump lance ses augmentations des droits de douane pour que le château de carte s'effondre. L'Ethereum a connu un gadin de 26 % et le Bitcoin y a aussi laissé quelques plumes avec une baisse de 4 %. Les memecoins, cryptomonnaies sans utilité économique surfant sur l'engouement autour d'une personnalité ou d'un phénomène viral sur internet, dégringolaient également. Dès le lendemain, après la pause décrétée par Donald Trump sur les droits de douane, le marché des cryptos retrouvait des couleurs. Ces mouvements de volatilité des actifs ne font que renforcer les ventes, alors que la valeur risque de remonter. Comparativement à d'autres placements, la situation démontre que les cryptomonnaies sont des actifs risqués à ne pas mettre entre toutes les mains. Sans compter que le marché manque de transparence.

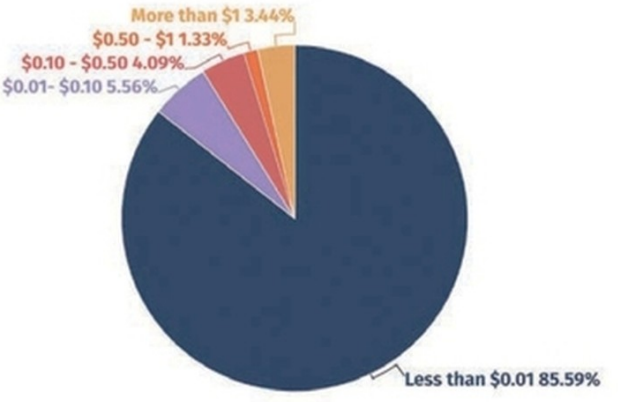
Autre signe, le patron de Microstrategy, grand acheteur de Bitcoin, vient d'interrompre ses achats. Marque de défiance ? Pas forcément, il attend juste de savoir qu'elle va être la stratégie de la Maison-Blanche vis-à-vis de ce type d'actifs, alors que tous les pays du monde commencent à vouloir reprendre la main pour encadrer ce marché. David Sacks devrait rapidement éclairer sa lanterne avec un objectif clair : *« Consolider la domination US dans l'espace numérique »*, avec un cadre régulatoire favorable et différents types de partenariats. La stratégie américaine autour des cryptomonnaies est ambitieuse, avec une réserve sur ces actifs dans le but de résorber les déficits de l'état sur les plus-values engendrées. On l'a vu, cette stratégie peut connaître des hauts mais aussi des très bas !

De plus, les règles du jeu ne seront pas égales partout. En Europe, la réglementation MiCA vise à une meilleure protection des utilisateurs, avec l'obligation d'une licence pour les prestataires de services de cryptomonnaies soumis à des règles strictes de transparence. Les cryptoactifs adossés à une monnaie comme le dollar devront désormais prouver qu'ils possèdent bien les réserves nécessaires pour garantir leur valeur. En outre, les plateformes crypto devront se conformer à des exigences plus strictes en matière de lutte contre le blanchiment d'argent et la fraude. De son côté, la Russie a permis le minage et autorise leur utilisation dans les paiements internationaux pour contourner les sanctions actuelles, mais augmente aussi son contrôle avec la création d'un registre obligatoire pour tout le matériel de minage. Il y va aussi d'une taxation sur les revenus qui s'échelonnent de 13 à 15 %. La Chine, comme d'autres pays, interdisent ce type d'actifs.

A savoir si la réserve prévue par Donald Trump connaîtra une destinée diverse de celle mise en place au Salvador, premier pays au monde à avoir mis en place ce mécanisme. Aujourd'hui, comme on dit dans le monde du spectacle, la réserve n'a pas trouvé son public, et le président, fan de Trump, revient un peu en arrière sur l'aspect stratégique de la cryptomonnaie.

Et en France ?

91 % des répondants connaissent les cryptoactifs, en légère hausse de 1 point par rapport à l'année précédente, et la moitié (50 %) affirment comprendre de quoi il s'agit. Environ 25 % des Français interrogés possèdent actuellement ou ont déjà acheté des cryptoactifs, un chiffre inférieur à la moyenne mondiale de 42 %. Les pays avec les taux de possession les plus élevés sont le Nigéria (73 %), l'Afrique du Sud (68 %), les Philippines (54 %), le Vietnam (54 %) et l'Inde (52 %). En outre, davantage de personnes en Asie et en Afrique prévoient d'investir dans les cryptoactifs au cours des 12 prochains mois. En Europe, bien que l'intérêt soit généralement plus faible, il a connu une hausse significative en France (27 %, +6 points) et en Allemagne (31 %, +10 points).



Part de marchés des cryptomonnaies selon leur prix.
Source: CoinMarketCap; Investing.comv

Dans l'Hexagone, les principaux obstacles à l'entrée dans l'écosystème des cryptoactifs restent la volatilité perçue du marché (22 %) et la prévalence des escroqueries (18 %). Bien que la perception de la volatilité ait diminué cette année, les inquiétudes persistent, en particulier dans des pays comme le Japon, la Corée du Sud et la Turquie. D'autres freins incluent le manque de connaissances sur le point de départ (13 %) et une incompréhension générale de l'objectif des cryptoactifs (9 %). Malgré ces barrières, les concepts associés aux cryptoactifs sont globalement perçus positivement : 10 % les voient comme *« une alternative à l'écosystème financier traditionnel »* (en baisse de 1 point), 8 % comme « l'avenir de l'argent », et 5 % comme *« l'avenir de la propriété numérique »* (en hausse de 1 point), bien que « la spéculation » demeure l'association la plus courante (17 %, en baisse de 1 point).

En juillet dernier, une étude d'Altindex recensait entre 8 500 et 10 000 cryptomonnaies. 96 % d'entre elles ont une valeur inférieure à 1 $. Seulement 131 dépassent les 50 cents. En fait, les cinq plus grosses cryptomonnaies représentent 82 % de la valeur du marché. Bref, pas de quoi s'extasier réellement alors que les jeux semblent faits et que la multiplication des memes n'y changera rien, si ce n'est des prurits de modes ou de tendances tout aussi temporaires que les couleurs lors d'une fashion week ! □

Datacenter

R&M enrichit ses solutions de monitorage

Le spécialiste helvétique de solutions d'infrastructure pour les réseaux de données et de communication enrichit ses solutions destinées au monitorage de réseaux sur la plateforme de répartition Netscale 48.

InteliPhy Analyzer complète la plateforme DCIM inteliPhy. C'est la deuxième génération du logiciel d'analyse. L'analyseur pilote et surveille jusqu'à 42 bandes de capteurs dans une baie au moyen d'un câblage bus. Il utilise le réseau de données local pour communiquer avec le système inteliPhy net pour la gestion centralisée des infrastructures, autorisant ainsi la surveillance à distance, en temps réel, des liaisons d'un réseau étendu.

Le nouvel inteliPhy Analyzer comprend un logiciel basé sur navigateur, offrant une visualisation claire des bandes de capteurs installées dans les baies. inteliPhy net, le logiciel centralisé DCIM (Data Center Infrastructure Management) de R&M, permet aux équipes techniques de voir d'un coup d'œil si toutes les connexions câblées fonctionnent et aboutissent au port souhaité. L'allocation des demandes d'entretien et de brassage est très aisée, tant dans le logiciel que sur l'afficheur de l'analyseur. Les lignes électriques et les états définis sont signalés au moyen de huit voyants LED. Il est possible d'entrer des commandes directement sur le clavier de l'analyseur.

Une interface bus à l'arrière sert à raccorder des capteurs et d'autres appareils. R&M a équipé l'inteliPhy Analyzer de deux prises pour une alimentation électrique redondante, ce qui augmente la sécurité d'exploitation du système de monitorage des réseaux de données dans des domaines d'application critiques, tels que la santé, l'approvisionnement énergétique ou la finance.

Il comprend un processeur Unix et des options logicielles. La mise à jour du firmware intégré se fait par clé USB. Une API Rest et le protocole MQTT soutiennent l'échange de données avec des applications externes, telles que le logiciel DCIM inteliPhy net.

Conformes au concept R&M des solutions d'infrastructure intégrées pour data centers, les produits inteliPhy s'installent selon les besoins dans les baies et les rangées de baies, mais aussi dans les systèmes DCIM d'autres fournisseurs. □

B.G

THD

Linkt lance une solution de connectivité en orbite basse pour les entreprises

Face aux défis posés par les zones isolées et mal desservies, l'opérateur télécom français Linkt lance une solution innovante de connectivité par satellite en orbite basse. Dédiée aux entreprises et aux collectivités, cette nouvelle offre s'inscrit dans une stratégie ambitieuse visant à fournir un accès au très haut débit partout en France, y compris dans les zones les plus reculées.

Dans un contexte où les usages numériques explosent, l'accès à une connexion fiable et performante est devenu indispensable pour les entreprises de toutes tailles. Certaines d'entre elles se situent cependant dans des zones dites « blanches » où il n'y a pas d'infrastructures réseau adéquates telles que la fibre optique ou une couverture 4G/5G suffisante. Filiale du groupe Altitude, l'opérateur télécom Linkt a décidé de répondre à cette problématique en développant une nouvelle solution de connectivité de pointe sans contrainte géographique basée sur une constellation de plus de 6 000 satellites en orbite basse. Cette technologie qui garantit des débits pouvant atteindre 220 Mb/s est capable d'assurer une connectivité performante, même dans les environnements territoriaux les plus reculés. *« Pour nos solutions satellites, nous collaborons avec un fournisseur, Marlink, qui est un revendeur agréé des solutions Starlink. Grâce à ce partenariat, nous proposons à nos clients des solutions satellites basées sur cette infrastructure, tout en ajoutant la valeur ajoutée propre à Linkt »*, détaille Louis Dreillard, chef de projet marketing chez Linkt.

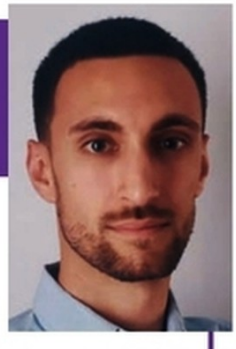
Une technologie prometteuse

La solution satellite de Linkt répond à des enjeux variés et stratégiques pour les entreprises. Elle constitue en effet une alternative rapide et fiable dans les zones dépourvues d'infrastructures terrestres. À l'instar des offres de connectivité en orbite basse introduites pas Starlink et OneWeb, elle est aussi plus simple à mettre en œuvre et plus fiable. *« Historiquement, l'Internet par satellite reposait sur des satellites géostationnaires situés à environ 35 000 km de la Terre. Bien que ces satellites soient toujours utilisés, les solutions en orbite basse présentent un avantage majeur : une latence beaucoup plus faible »*, précise le spécialiste de Linkt. *« Avec les satellites géostationnaires, la latence est d'environ 700 millisecondes. En revanche, avec les satellites en orbite basse, nous atteignons une latence inférieure à 100 millisecondes, souvent entre 35 et 40 millisecondes. Cette réduction de la latence est une innovation clé qui attire de plus en plus d'entreprises vers les solutions satellites, notamment lorsqu'il n'existe pas d'infrastructure terrestre viable. »*

Un atout stratégique

Ce nouveau type d'offre favorise le développement des activités, la modernisation des outils numériques et la transition digitale. En complément des réseaux terrestres, tels que la fibre, le satellite proposé par Linkt joue un rôle de filet de sécurité en cas de panne, garantissant une connexion résiliente et continue pour maintenir les opérations critiques des entreprises. Alors que le réseau cuivre touche à sa fin, la solution satellite de Linkt se positionne en outre comme une alternative stratégique pour les entreprises situées dans des zones où la fibre n'est pas encore disponible. Conçue pour répondre aux attentes variées des entreprises, elle offre des performances optimisées en matière de débit et de latence, et prend en charge des technologies avancées telles que MPLS, SDWAN et SASE.

Louis Dreillard,
chef de projet marketing chez Linkt



« Le satellite est désormais pertinent pour plusieurs raisons. D'un point de vue tarifaire, les coûts ont considérablement baissé par rapport à il y a dix ans, et les performances, notamment en termes de débit, se sont aussi beaucoup améliorées »

Par rapport à l'offre de Starlink, initialement conçue pour les particuliers avec des abonnements à bas coût, Louis Dreillard souligne que Linkt se distingue en proposant des fonctionnalités spécifiquement adaptées aux besoins des entreprises : *« Notre différenciation repose sur la capacité à offrir une solution complète et adaptée aux entreprises, avec une qualité de service supérieure à ce que proposent d'autres acteurs. Link offre une IP fixe garantie, des services supplémentaires comme un VPN IP pour interconnecter des sites, ainsi que des outils de supervision pour assurer la qualité de l'accès Internet. Nous bénéficions également d'une expertise télécom reconnue, que nous avons accumulée avec d'autres technologies (fibre dédiée ou mutualisée) que nous appliquons désormais au satellite. »*

Connexion en orbite basse : les défis de l'installation

L'installation de ces solutions pose des enjeux techniques et opérationnels spécifiques. De l'audit préalable à la configuration des équipements, chaque étape est cruciale pour garantir une performance optimale et une intégration fluide dans l'écosystème numérique des entreprises. Pour Louis Dreillard, il est indispensable d'effectuer un audit préalable pour garantir des services à forte valeur ajoutée : *« Théoriquement, un site est éligible dès lors qu'il dispose d'une vue dégagée vers le ciel, sans obstacle pouvant obstruer la visée des satellites. Chez Linkt, nous réalisons une prestation d'audit satellite en amont pour permettre de vérifier, d'une part, si le site est conforme à l'utilisation de cette technologie, et d'autre part, déterminer un emplacement optimal pour l'antenne. Une fois l'audit validé, nous procédons à l'installation avec nos techniciens qui posent l'antenne, effectuent les câblages nécessaires et installent notre propre routeur (nous n'utilisons pas le routeur standard de Marlink). Cette démarche garantit une qualité de service optimale et permet d'ajouter des services à valeur ajoutée. »*

De nouvelles fonctionnalités pour 2025

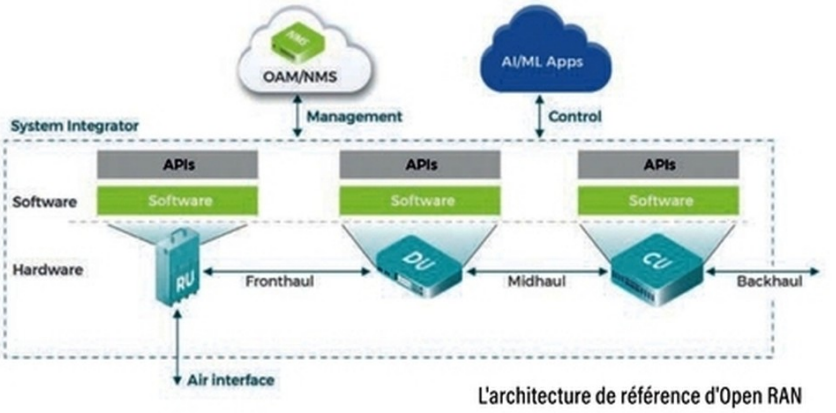
La solution de connectivité par satellite de Linkt, qui a déjà séduit plusieurs de ses clients, va s'enrichir de nouvelles fonctionnalités au cours de l'année 2025. L'opérateur prévoit notamment de proposer de nouveaux types d'abonnements. *« Pour répondre aux besoins de nos clients, nous prévoyons de commercialiser prochainement des abonnements mobiles avec des antennes capables de fonctionner en mouvement comme sur les véhicules des pompiers. Certains véhicules sont dédiés à la communication pendant les interventions (comme les incendies) et ils pourront être équipés d'antennes satellites leur permettant d'accéder au très haut débit, en mouvement et dans n'importe quelle zone géographique. Nous envisageons également de proposer des abonnements itinérants avec, cette fois-ci, des antennes qui pourraient être déplacées d'un site à un autre pour répondre à des besoins ponctuels. Une entreprise pourrait par exemple déployer temporairement une antenne itinérante sur un site spécifique. Que cela soit pour les usages mobiles ou itinérants, nous utiliserons principalement des antennes résilientes à la fois plus robustes et légèrement plus grandes. Elles demeureront néanmoins compactes et adaptées à un usage mobile. »* ☐ **J.C**

Technologies

Les prédictions de Juniper pour 2025

Alors que son rachat par HPE est en discussion par les autorités de marché américaines, Juniper Networks voit la mise en œuvre de plusieurs technologies pour l'année à venir, et cela n'est pas forcément de l'intelligence artificielle !

Première tendance technologique relevée par la société, le déploiement de l'Open RAN (O-RAN) devrait s'accélérer, notamment chez les opérateurs historiques, et ce aussi bien en Europe qu'en Amérique du Nord. Cette tendance sera notamment portée par de nouvelles réglementations, mais aussi par la demande grandissante du marché. L'automatisation O-RAN, en particulier pour les applications RIC en temps non réel (Non-RT RIC) et quasi temps réel (Near-RT RIC), améliore la qualité de service, optimise les performances réseau et réduit la consommation d'énergie. L'évolution des normes O-RAN et l'enrichissement de son écosystème permettront également l'essor d'applications collaboratives basées sur l'IA, sur des réseaux multifournisseurs.



L'architecture de référence d'Open RAN

Le cantique du quantique

L'informatique quantique, sujet d'intérêt et d'inquiétude, attire des investissements massifs, estimés à plus de 100 milliards en 2024. Face à la future menace pour la sécurité des réseaux, les États examinent déjà des solutions de sécurité quantique, notamment la cryptographie post-quantique (PQC). Les organisations des secteurs (finance, santé, administrations publiques) ayant des exigences de sécurité élevées sont-elles à la demande de recherche hybride ? Comme l'ordinateur et l'internet quantiques ne seront pas disponibles en 2025, il sera possible d'assister à la première intrusion dans un réseau à sécurité quantique, sous l'impulsion de cybercriminels à la recherche de notoriété. Afin de contrecarrer ces risques, des progrès significatifs dans l'adoption et le déploiement de solutions de sécurité quantique seront visibles dans les prochains mois.

Des réseaux plus performants…

Les réseaux urbains et périphériques subiront une transformation majeure, poussés par la 5G, les besoins en haut débit et faible latence, les plateformes de vidéo en continu et bien entendu les services pilotés par l'IA. La demande de 100G et 400G, rendue possible par les optiques ZR, contribuera au renouvellement de ces infrastructures pour tendre vers des architectures distribuées intégrant routage, switching et sécurité. L'automatisation basée sur l'IA permettra quant à elle une gestion proactive du réseau et une orchestration de bout en bout.

Avec l'AIops…

L'automatisation basée sur les intentions, stimulée par des AIOps considérablement perfectionnés, dont des modèles d'inférence de l'IA, l'IA générative et les jumeaux numériques, transformera la gestion des réseaux. L'analyse prédictive améliorera la compréhension du réseau et réduira les temps de résolution des incidents. L'IA générative simplifiera les tâches des opérateurs, tandis que les jumeaux numériques permettront une expérimentation sans risque. Cette évolution poussera les opérateurs à privilégier les investissements axés sur les résultats plus que sur la technologie, et alignés sur leurs objectifs stratégiques.

Et toujours plus verts !

L'IA et l'automatisation joueront un rôle clé dans l'optimisation des réseaux et la réduction de la consommation d'énergie, notamment via la mise en œuvre de modes de veille. Enfin, l'IA permettra également une meilleure observabilité des émissions carbone et de la dissipation thermique, grâce à des outils de suivi et des traqueurs de durabilité, offrant ainsi une meilleure compréhension globale de l'empreinte carbone et les axes d'amélioration. □

B.G

Concurrence

Avec sa levée de fonds record, Databricks met la pression sur Snowflake

Dans la lutte acharnée que se livrent les deux data platforms, la dynamique semble actuellement du côté de Databricks, qui vient de réaliser la plus grosse levée de fonds de l'histoire de la tech. Mais Snowflake a aussi de nombreux arguments à faire valoir.

L'annonce a sonné comme un coup de tonnerre dans le monde des plateformes de données. Le 17 décembre 2024, Databricks réalisait la plus grosse levée de fonds de l'histoire de la tech, avec pas moins de dix milliards de dollars récoltés auprès d'un panel de fonds en capital-risque de premier plan. La data platform spécialisée dans l'IA et les traitements des masses de données coiffe ainsi au poteau OpenAI, qui avait déjà réalisé une belle levée de fonds de 6,6 milliards de dollars début octobre 2024. Parmi ceux qui ont mis la main au panier, on compte notamment Thrive Capital (également présent dans la levée de fonds d'OpenAI, et connu pour avoir très tôt débusqué des perles comme GitHub, RobinHood, Slack et Spotify), l'incontournable Andreessen Horowitz, DST Global, le fonds souverain singapourien GIC, ainsi que quelques autres.

Cette treizième levée de fonds porte la valorisation estimée de Databricks à 62 milliards de dollars, en hausse de 50 % sur un an, devant sa rivale Snowflake, qui est, elle, cotée en bourse et dont la valorisation actuelle est d'un peu moins de 55 milliards.

Une rivalité qui remonte à loin

Databricks et Snowflake constituent aujourd'hui les deux poids lourds du marché des plateformes de données. Toutes deux sont nées avec l'essor du cloud. Elles viennent toutefois de deux mondes différents.

Snowflake naît en 2012, alors qu'une vague d'entreprises commençait à migrer ses données vers un, voire plusieurs clouds. S'impose alors la nécessité pour ces professionnels de gérer de larges quantités de données dans différents environnements, avec le risque que des morceaux ou « silos » de données stockés quelque part n'interagissent plus avec le reste des systèmes de l'entreprise. D'où la nécessité de solutions pour gérer holistiquement ces données.

C'est ce que propose Snowflake, à travers un entrepôt de données full cloud, avec une solide couche SQL et une prise en main aisée pour les professionnels de la business intelligence.

Databricks, lancé un an plus tard, se caractérise plutôt par une solution centrée sur le traitement des masses de données via le framework big data open source Apache Spark, dont les fondateurs de Databricks sont à l'origine. La jeune pousse évolue ensuite rapidement vers le traitement de workloads d'apprentissage automatique, pour laquelle sa technologie est parfaitement adaptée, avec notamment la librairie MLlib. En 2020, elle conceptualise également le lakehouse, convergence entre lacs (data lake) et entrepôts de données (data warehouse), qui permet de stocker, gérer et analyser tous types de données (structurées, semi-structurées et non structurées) au même endroit, dans un format unique et open source, baptisé Delta.

« Les entreprises qui travaillent beaucoup sur des entrepôts, comme Oracle, ou même des technologies un peu plus spécialisées, comme Vertica ou Teradata, vont très bien se retrouver dans Snowflake, parce que c'est très orienté SQL et SQL Analytics, et très puissant dans ce domaine-là. Par contre, ceux qui viennent du monde du big data, qui ont eu très tôt l'impératif de traiter des

Thomas Dallemagne,
Partner Advisory
Micropole.

« On va même pouvoir concevoir les interfaces utilisateurs sur les outils d'IA complètement à l'intérieur de Snowflake. Databricks s'efforce de progresser sur ce terrain-là en proposant des solutions plus faciles à prendre en main »

gros volumes transactionnels avec des latences assez faibles, et l'habitude de fonctionner avec des outils très techniques, vont plutôt s'orienter vers Databricks », note Thomas Dallemagne, du cabinet de conseil Micropole spécialisé sur les données et l'IA.

Snowflake accélère sur l'IA, Databricks se lance dans les tables open-source

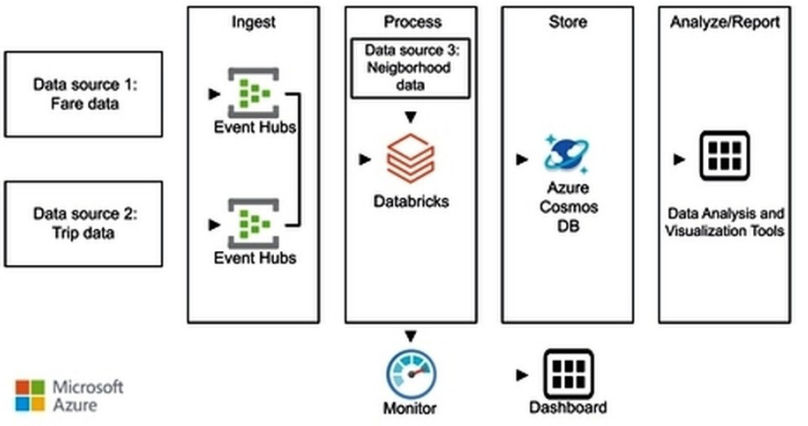
Chacun vient désormais chasser sur les terres de l'autre. Snowflake s'efforce de développer son offre autour de l'apprentissage automatique pour surfer sur la vague de l'IA. Elle a par exemple tissé un partenariat avec Nvidia, maître incontesté des cartes graphiques (GPU) nécessaires pour entraîner les modèles d'IA générative. Elle a également sorti Cortex AI, son propre grand modèle linguistique (LLM). Dans sa rivalité avec Databricks, Snowflake peut compter sur une place de marché des données plus riche que celle de sa rivale, avec son data cloud conçu pour exécuter, distribuer et monétiser des applications basées sur les données, les cadres de l'entreprise répétant à l'envi que *« pour avoir une stratégie autour de l'IA, il faut d'abord avoir une stratégie autour des données »*.

Databricks, de son côté, vient également titiller Snowflake via une stratégie d'acquisitions tous azimuts. En juin 2024, le rachat de Tabular, pour lequel l'entreprise a déboursé un milliard de dollars, lui permet de se positionner avantageusement sur Apache Iceberg, crucial pour la gestion des données cloud, et de concurrencer Snowflake sur ses points fort, l'interopérabilité et la gestion des données. En juin 2023, elle se positionne également sur les LLM avec le rachat de MosaicML pour 1,3 milliard de dollars.

Databricks va-t-elle entrer en bourse ?

Si les deux sociétés ont les reins solides, la dynamique semble actuellement plutôt en faveur de Databricks, qui conserve un avantage sur l'IA, grâce à sa spécialisation historique et à ses acquisitions. Sa dernière levée de fonds gargantuesque va en outre lui permettre de réaliser de nouvelles acquisitions stratégiques, d'accélérer dans l'IA et de pousser son développement à l'international.

La grande question est désormais de savoir si Databricks va rejoindre sa rivale en entrant en bourse ou non. Une IPO avait été envisagée en 2021/2022, avant d'être remise aux calendes grecques face au contexte devenu très défavorable pour les entrées en bourse, avec l'augmentation brutale des taux d'intérêt pour lutter contre l'inflation. Alors que le contexte montre des signes d'amélioration, Ali Ghodsi a évoqué lors de l'annonce de la levée de fonds une possible entrée en bourse en 2025, sans toutefois s'engager fermement sur la question.



Le traitement des flux dans Databricks

La capacité dont fait preuve l'entreprise à se financer (très) confortablement sur le marché primaire écarte pour l'heure la nécessité d'une entrée en bourse rapide, d'autant que le fait de rester privée a ses avantages. Databricks n'a pas à divulguer autant d'informations à ses concurrents que si elle était publique, et n'est pas soumise aux performances trimestrielles qui nécessitent de répondre aux exigences des marchés financiers. Elle peut, par exemple, se permettre de dépenser davantage à court terme pour se construire un avantage décisif dans le futur, sur des technologies de pointe comme les grands modèles de langage. Databricks est loin d'être la seule à reculer son entrée en bourse. Thomas Laffont, de Coatue, un investisseur de la Silicon Valley, estimait en septembre 2024 dans une conférence, que l'on comptait actuellement 1 440 sociétés non cotées aux États-Unis valorisées à plus d'un milliard de dollars, un chiffre énorme qui illustre la frilosité des entrepreneurs à se frotter à Wall Street.

Pourquoi Snowflake n'a pas dit son dernier mot

Malgré la dynamique positive dont jouit Databricks, Snowflake compte encore de nombreux atouts dans son escarcelle. *« À mon sens, la principale proposition de valeur de Snowflake, c'est l'ergonomie. Pour un analyste, pour un ingénieur des données, c'est une interface très léchée, très facile à configurer, à passer à l'échelle, on crée ses entrepôts de données virtuels, on peut les taguer, les assigner, etc., ça fonctionne très bien. Or, Cortex suit la même philosophie, en se profilant comme une boîte à outils. On va même pouvoir concevoir les interfaces utilisateurs sur les outils d'IA complètement à l'intérieur de Snowflake. Databricks s'efforce de progresser sur ce terrain-là en proposant des solutions plus faciles à prendre en main, mais ils ne sont pas encore au même niveau »*, estime Thomas Dallemagne.

Michelle Swan, de l'investisseur Tercera, voit pour sa part dans le programme partenaire de Snowflake un avantage sur son concurrent. *« Snowflake possède un programme de partenaires très structuré où chacun peut facilement se différencier. »* □

G.R

Cloud

« Les communautés open source se sont emparées du sujet de l'IA »

Avec Rémy Mandon, country manager France de Red Hat, retour sur la popularité croissante de l'open source dans les entreprises, ainsi que sur les synergies entre ce paradigme et l'IA.

Plein cap sur l'IA et le cloud hybride ! Les résultats d'IBM au troisième trimestre indiquent une forte croissance de la division logicielle de l'entreprise, qui affiche près de 10 % de croissance d'une année sur l'autre, là où la division infrastructure, le cœur de métier historique de l'entreprise, affiche de son côté un recul de 7 %. *« Nos investissements dans le logiciel, dans le cadre du repositionnement de notre portfolio au cours des dernières années, s'avèrent payants. Au troisième trimestre, le logiciel a constitué un large relais de croissance et compte désormais pour 45 % de notre chiffre d'affaires »*, s'est félicité James Kavanaugh, vice-président et directeur financier d'IBM, lors de l'annonce des résultats.

Une performance notamment imputable à Red Hat, spécialiste de l'open source racheté par IBM en 2019, qui affiche 14 % de croissance. Plus grosse acquisition jamais réalisée par le géant de l'informatique, qui avait déboursé 34 milliards dans l'opération, le rachat de Red Hat constituait alors un pari audacieux de la part d'IBM. La société, qui fait partie des barbes grises de l'IT, s'efforçait de se réinventer en se focalisant sur l'open source et le logiciel afin de trouver de nouveaux relais de croissance.

Un pari qu'IBM a depuis poursuivi par d'autres acquisitions, comme celle, début novembre, de Neural Magic, concepteur d'algorithmes taillés pour accélérer les charges d'inférence d'IA générative, par Red Hat. Ainsi que par des partenariats, comme celui signé début décembre entre Red Hat et Amazon, afin d'étendre la disponibilité des solutions open source de Red Hat dans la Marketplace d'AWS. Une évolution qui doit notamment permettre à Red Hat de mieux aider ses clients à migrer leurs machines virtuelles et charges de travail conteneurisées vers le cloud, dans le but de déployer plus facilement des applications autour de l'IA.

Nous nous sommes entretenus avec Rémy Mandon, country manager France de Red Hat, pour évoquer avec lui le rôle grandissant des solutions open source dans la conduite de la transformation des entreprises. □

RÉMY MANDON, COUNTRY MANAGER FRANCE DE RED HAT

L'informaticien : Qu'est-ce qui explique, selon vous, la popularité de l'open source aujourd'hui dans les entreprises ? En quoi ce paradigme est-il adapté aux évolutions technologiques récentes ?
Rémy Mandon : Un premier élément me semble être la capacité de retrouver de l'autonomie stratégique grâce à la portabilité permise par l'open source. Un client qui choisit Red Hat Enterprise Linux ou OpenShift peut faire tourner ses applications sur l'environnement de son choix, sur ses propres serveurs ou dans le cloud, sans changer le code applicatif.

Cette portabilité permet de donner des options aux clients. Si le hardware devient trop cher, si le fournisseur cloud augmente trop ses tarifs, la réversibilité est beaucoup plus simple avec une solution open source. Or, on a vu récemment, avec le sujet Broadcom/VMWare, combien les dépendances commerciales ou techniques pouvaient poser problème.

Le deuxième avantage de l'open source, c'est que les compétences nécessaires pour se servir des différentes solutions sont facilement disponibles sur le marché. Tandis que pour les solutions défendues par de la propriété intellectuelle, c'est déjà plus compliqué. Dans l'open source, il y a des consultants et des intégrateurs (de taille mondiale ou régionale), des experts qui travaillent en freelance, des entreprises qui font du développement open source leur ADN et qui ne font que ça. Donc sur le marché, les compétences sont quand même beaucoup plus facilement disponibles. C'est quelque chose que les professionnels apprécient.

À l'heure où toutes les entreprises cherchent à optimiser leur fonctionnement grâce à l'IA, dans quelle mesure le choix de l'open source est-il pertinent pour cette technologie ?
Aujourd'hui, lorsqu'on veut déployer des solutions autour de l'IA générative notamment, deux options s'offrent à nous. La première consiste à entraîner ses propres algorithmes sur ses infrastructures. Cette solution n'est cependant pas celle qui convient à la majorité des entreprises. En effet, il faut alors revoir ses infrastructures, et notamment intégrer des GPUs, qui d'une part coûtent très cher, et d'autre part demandent souvent des délais de livraison conséquents, la demande étant très supérieure à l'offre.

La plupart des entreprises penchent donc pour une deuxième option, à savoir accueillir les algorithmes déjà entraînés (généralement par leur fournisseur cloud) et faire ensuite de l'inférence, c'est-à-dire adapter ces algorithmes à ses besoins. Faire tourner des modèles avec un ratio coût/performance raisonnable est alors tout à fait possible. Il existe des grands modèles de langage pour lesquels les infrastructures actuelles des clients peuvent permettre de faire de l'inférence sans avoir à investir massivement dans de nouvelles architectures coûteuses.

Or, les communautés open source se sont massivement emparées de ce sujet, là où les éditeurs propriétaires se concentrent plutôt sur la qualité de leur algorithme, sur l'entraînement lui-même et l'introduction de milliards de paramètres. Concrètement, cela signifie que sur les 80 milliards de paramètres que comprend un modèle propriétaire, on en garde entre 3 et 10 milliards, dans lesquels on a enlevé les Haïkus japonais, le serbo-croate, le Swahili, etc., autant d'éléments qui ne me servent pas beaucoup si je suis une grande banque française qui cherche à mettre en place des cas d'usage autour de l'IA, par exemple.

C'est donc aussi un élément de différenciation pour Red Hat ?
Effectivement. C'est pour ça que Red Hat et IBM ont mis en open source un modèle qui s'appelle Granit, co-développé par IBM Research et Red Hat. Il a été nettoyé de tous les propos haineux ou protégés par de la propriété intellectuelle, et conserve une performance comparable aux très gros modèles de langage. En fonction de ses besoins, le client peut choisir un nombre de paramètres qui va de 8 à 11 milliards, et se retrouve ainsi avec un modèle plus facile à entraîner, qui coûte moins cher et requiert moins de puissance de CPU ou GPU.

C'est aussi pour cela que Red Hat a racheté NeuralMagic, une solution d'optimisation d'inférence, qui était déjà utilisée dans OpenShift, et dont le code est open source. Ce rachat nous permet de garantir la pérennité de ce code-là, qui est extrêmement efficace pour optimiser les algorithmes à exécuter sur des environnements existants.

On voit aujourd'hui émerger des algorithmes et des grands modèles de langage spécifiques dans tous les domaines, du juridique aux chatbots pour aider les clients ou faire du service après-vente, en passant par la comptabilité ou même le fait de développer du code. Aujourd'hui, les métiers fourmillent d'idées, et les directeurs informatiques se disent *« Et moi, comment je fais pour faire tourner tout ça chez moi, pour éviter que mes coûts explosent, et pour que mes équipes aient les bonnes compétences ? »* C'est vraiment là que l'open source prend tout son sens.

À travers Red Hat, IBM s'efforce également de faire rimer cloud hybride et open source. Pourquoi ce choix ?
Il y a encore quelques années, le consensus était que tout le monde allait migrer vers le cloud à plus ou moins brève échéance. Or, on s'est depuis rendu compte que ce n'est pas aussi simple que cela. Certains ont des contraintes réglementaires, d'autres des contraintes techniques. Dans la banque, pas exemple, tout est très réglementé. Dans l'armée, il faut des services déconnectés avec des environnements qui ne peuvent pas être dans le cloud. Au-delà de ces exemples, de plus en plus d'entreprises se rendent compte que la dépendance au fournisseur cloud constitue un vrai risque industriel.

Dans ce contexte, le cloud hybride s'impose donc de plus en plus. L'open source est très adapté à ce paradigme, pour les questions de portabilité que nous évoquions plus haut.

Sur quel autre axe stratégique faites-vous le pari pour l'avenir ?
La question de la virtualisation va à mon sens devenir de plus en plus stratégique. Ce qui s'est passé avec VMWare et Broadcom a fait prendre conscience aux clients de leur dépendance commerciale et technique et des risques que cela impliquait. Maintenant, il y a ainsi une vraie réflexion de fond, non plus sur les aspects commerciaux, mais sur la technologie de virtualisation elle-même. Finalement, est-ce qu'un client ne pourrait pas simplifier drastiquement son infrastructure et faire tourner des containers (et donc les applications et algorithmes applicatifs) directement sur du bare-metal ?

Nous pensons que le domaine de la virtualisation va beaucoup changer et continuer à tracter en avant la containérisation. □

G.R

Moteur de recherche

En 2025, Perplexity va-t-elle rattraper Google ?

La jeune pousse dispose d'une technologie prometteuse et d'une certaine popularité parmi les primo adoptants de la Silicon Valley. Se pose toutefois la question de son modèle économique et de sa dépendance aux LLM d'entreprises tierces.

La domination de Google sur la recherche en ligne a longtemps semblé inébranlable. La sortie de ChatGPT fin 2022, l'investissement massif de Microsoft dans OpenAI et l'annonce de l'intégration du chatbot à Bing ont certes momentanément soulevé l'idée d'un danger pour Google. Mais la ferveur est rapidement retombée et la part de marché du géant de la recherche en ligne est demeurée inchangée[1].

Cependant, depuis quelques mois, une petite musique résonne dans la Silicon Valley : Google ferait enfin face à un adversaire à sa mesure, sous la forme d'une jeune pousse baptisée Perplexity AI. Jensen Huang, le patron de Nvidia, affirme l'utiliser tous les jours, et Tobi Lutke, dirigeant de Shopify, a également vanté les mérites de l'application sur son compte X. La tendance n'est pas cantonnée à l'autre côté de l'Atlantique. *« Depuis que j'ai découvert Perplexity, je n'utilise quasiment plus Google »*, nous confie ainsi un cadre au sein d'une grande entreprise française des télécoms.

La jeune pousse est soutenue par de nombreux investisseurs comptant parmi les plus prestigieux, de Jeff Bezos à Yann LeCun (à la tête de l'IA chez Meta) en passant par Nvidia, Softbank et Sequoia Capital.

Les atouts de Perplexity

Le moteur de recherche de Perplexity AI repose sur l'usage de plusieurs grands modèles de langage : le sien, développé en interne, mais aussi Claude (Anthropic), GPT-4 (OpenAI) ou encore Llama (Meta), en sachant qu'une version premium facturée 20 euros par mois permet d'accéder à davantage de modèles que la version gratuite. À chaque requête, ces algorithmes sondent le web pour fournir en quelques secondes à l'utilisateur des réponses sous forme de textes synthétiques et argumentés, assortis des différents liens utilisés en guise de source. L'entreprise fait de la fiabilité et de la précision de son chatbot, l'un de ses principaux arguments de vente, et s'il lui arrive parfois d'halluciner, force est de constater que ses réponses sont majoritairement de très grande qualité.

Passé par OpenAI et Google DeepMind, Aravind Srinivas, le dirigeant de Perplexity AI, est convaincu du fait que l'IA générative incarne la nouvelle manière d'accéder à l'information sur la toile, et d'avoir développé le produit idéal pour y parvenir. Il a notamment mis l'accent[2] sur la faible latence des réponses fournies par son chatbot, et la compression maximale des ressources informatiques nécessaires pour fournir ces réponses, de façon à passer à l'échelle plus facilement en limitant les coûts. La jeune pousse est également parvenue à concevoir une solution qui s'appuie sur les meilleurs LLM disponibles pour fournir une qualité de réponse optimale.

Les temps sont durs pour Google

Si la jeune pousse peut compter sur ses propres atouts pour concurrencer Google, elle arrive également à un moment où le géant de la recherche en ligne est particulièrement vulnérable. Nombreuses sont les jeunes pousses à avoir voulu remettre en cause le monopole de Google au fil des années : l'une des plus prometteuses, Neeva, a jeté l'éponge en 2023. Cependant, un consensus émerge actuellement autour du fait que la qualité des recherches Google a diminué. Une étude allemande parue l'an passé montre par exemple que celle-ci est plombée par le spam de contenus SEO, contre lesquels Google semble avoir de plus en plus de mal à lutter[3].

L'entreprise californienne est également la cible de plusieurs procès, aux États-Unis, au Royaume-Uni et dans l'UE, qui l'accusent d'étouffer la concurrence en profitant de sa position dominante. Une cour fédérale américaine a notamment décrété l'été dernier que Google exerçait un monopole sur

1 https://gs.statcounter.com/search-engine-market-share.

2 https://www.youtube.com/watch?v=jksGQhMtXjo

3 https://mashable.com/article/google-search-low-quality-research?ref=hackernoon.com

L'interface de Perplexity

la recherche en ligne, violant la Section 2 du Sherman Antitrust Act de 1890. L'entreprise pourrait en conséquence être contrainte de se séparer de son navigateur Google Chrome, ce qui affaiblirait considérablement la richesse et la cohérence de l'écosystème dont Google tire sa domination, ainsi que sa puissante machine à cash. Contrairement aux requêtes effectuées sur des navigateurs rivaux comme Safari ou Firefox, celles effectuées sur Chrome permettent en effet à Google de collecter un grand nombre de données supplémentaires au-delà de la recherche en elle-même, comme les recherches annexes effectuées par l'utilisateur, sa localisation, la façon dont il réagit à certaines publicités plutôt qu'à d'autres, ainsi que ses sites favoris. Tout cela est mis au service de la capacité de Google à effectuer des publicités ciblées d'une grande précision.
Dans ce contexte, Perplexity dispose donc d'une fenêtre de tir idéale pour venir concurrencer le mastodonte de la recherche en ligne.

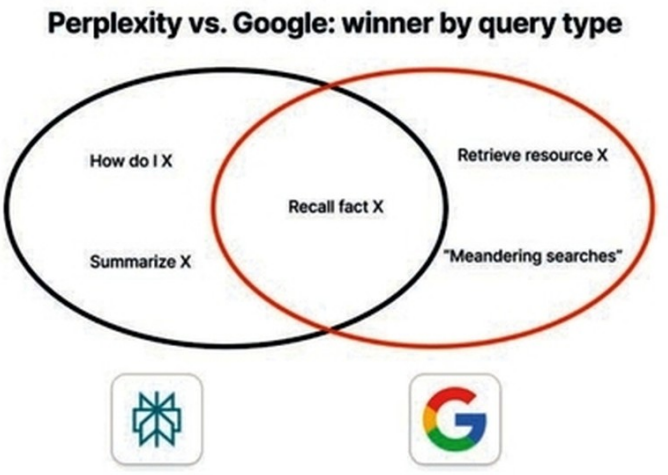
Le dilemme de l'innovateur

Si Perplexity est incontestablement un excellent produit, la jeune pousse est encore très loin d'avoir remporté la bataille contre Google, et doit s'attendre à de nombreuses difficultés sur sa route. Les premières tiennent aux forces dont dispose Google, l'une des entreprises les plus puissantes au monde.

Celle-ci dispose d'abord d'un puissant effet réseau : 8,5 milliards de requêtes[4] lui sont soumises chaque jour. Face à cela, les 100 millions de requêtes hebdomadaires de Perplexity font figure d'une goutte d'eau dans l'océan. Changer les habitudes de milliards d'utilisateurs s'annonce difficile. *« On dit que l'IA générative est le nouveau concurrent de Google. Moi, ça ne me semble pas du tout évident. D'ailleurs, le nombre des recherches Google ne décroît pas, le moteur de recherche est toujours utilisé massivement »*, estime Gilles Moyse, docteur en IA et fondateur de la jeune pousse Récital.

Google dispose également de solides compétences autour de l'IA, avec son propre LLM (Gemini) et son assistant basé sur l'IA générative (Bard). L'entreprise a également ajouté une fonction d'IA générative à son moteur de recherche, qui propose des réponses synthétiques sur le modèle de celles de Perplexity au-dessus des liens qui s'affichent traditionnellement après une requête. Google compte certains des meilleurs ingénieurs en IA au monde parmi ses rangs, et dispose d'un confortable matelas de cash pour innover. Grâce à son effet de réseau, l'entreprise peut en outre rapidement déployer ses solutions auprès d'un très grand nombre d'utilisateurs, afin de les tester et les améliorer.

Avec un bémol, toutefois : Google fait face au dilemme de l'innovateur, conceptualisé par Clayton Christensen dans son célèbre livre de 1997. Si elle reste les bras croisés, elle risque de perdre la course à l'innovation. Mais si elle vient chasser sur les terres de Perplexity, elle risque de remettre en cause l'intégralité de son modèle d'affaires, qui repose sur la publicité, et nécessite



Les points forts des différents moteurs de recherche

que les internautes cliquent sur les liens que Google propose à chaque recherche.

Perplexity doit encore trouver son modèle économique

Un dilemme qui souligne cependant le fait que l'IA générative n'a pas encore trouvé son modèle d'affaires, et constitue donc également un problème pour Perplexity. Il est en effet douteux[5] que l'entreprise puisse générer des revenus suffisants avec sa formule payante : Sam Altman a récemment admis que la version Pro de ChatGPT, pourtant facturée 200 dollars par mois, perdait elle-même de l'argent.

Une autre faiblesse de Perplexity constitue le revers de l'un de ses points forts : en s'appuyant sur plusieurs LLM différents développés par des entreprises tierces, l'entreprise a conçu un agrégateur capable de fournir des réponses très précises. Mais elle se met également à la merci de ces entreprises, qui peuvent à tout moment choisir de lui couper l'accès à leurs solutions.

Enfin, au même titre qu'OpenAI, Perplexity fait face à une fronde de la part des éditeurs et créateurs de contenus, qui lui reprochent d'utiliser leurs textes sans leur autorisation pour enrichir son produit. L'entreprise est notamment accusée d'utiliser des articles qui se trouvent pourtant derrière un verrou d'accès payant[6] et d'avoir plagié des travaux journalistiques[7]. En réaction, Perplexity a mis en place[8] un modèle de partage des revenus avec les éditeurs dont elle utilise les contenus, ce qui relance toutefois la question de son modèle d'affaires… Google, de son côté, réalise près de 74 milliards de dollars de bénéfice par an. □

4 https://seo.ai/blog/how-many-people-use-google

5 https://productify.substack.com/p/how-perplexitys-growth-is-changing

6 https://www.threads.net/@vthallam/post/C59LmZ0A0uK?hl=en&ref=hackernoon.com

7 https://www.wired.com/story/perplexity-plagiarized-our-story-about-how-perplexity-is-a-bullshit-machine/?ref=hackernoon.com

8 https://www.cnbc.com/2024/07/30/perplexity-ai-to-share-revenue-with-publishers-after-plagiarism-accusations.html?ref=hackernoon.com

IA Gemini débarque sur Google Workspace en version française !

Google Cloud annonce une avancée majeure pour les utilisateurs de Google Workspace, avec le lancement de Gemini en français et six autres langues. Désormais accessible directement depuis Gmail, Google Docs, Sheets, et Drive, Gemini vise à démocratiser l'utilisation de l'intelligence artificielle générative dans le quotidien des professionnels. Cette intégration accélère l'adoption de l'IA dans les outils collaboratifs avec l'objectif d'améliorer la productivité et la qualité de travail.

La compétition entre les géants de l'IA s'intensifie dans le secteur des suites bureautiques. Google Cloud a franchi une étape majeure dans l'évolution de ses outils collaboratifs en annonçant l'arrivée de son intelligence artificielle générative Gemini au sein de Google Workspace. Depuis le 21 novembre 2024, cette fonctionnalité est désormais accessible en français et dans six autres langues : allemand, coréen, espagnol, italien, japonais et portugais. Cette avancée illustre la volonté de Google de démocratiser l'IA et de transformer les pratiques professionnelles. Les premières applications IA pour Workspace ciblent principalement les fonctions marketing, commerciales et de relation client. Ces secteurs pourront tirer parti des capacités avancées de résumé, d'analyse et de génération de contenu offertes par Gemini. En exploitant les données issues des e-mails, documents et fichiers dans un environnement sécurisé et conforme à la gouvernance des données, les entreprises peuvent accroître leur efficacité et leur créativité. Pour s'en rendre compte, Google a mené une étude auprès de 3 200 utilisateurs anglophones utilisant déjà Gemini pour Workspace. *« Les utilisateurs gagnent en moyenne 105 minutes par semaine, et 75 % d'entre eux estiment que cela améliore significativement la qualité de leur production. Le service leur permet de réaliser des documents mieux présentés, mieux organisés et avec plus d'impact. Il simplifie également leurs tâches répétitives et fastidieuses, et leur laisse de fait plus de temps pour des tâches à forte valeur ajoutée »*, détaille Frédéric Arnoux, chief technical officer Europe Middle East & Africa de Google Workspace.

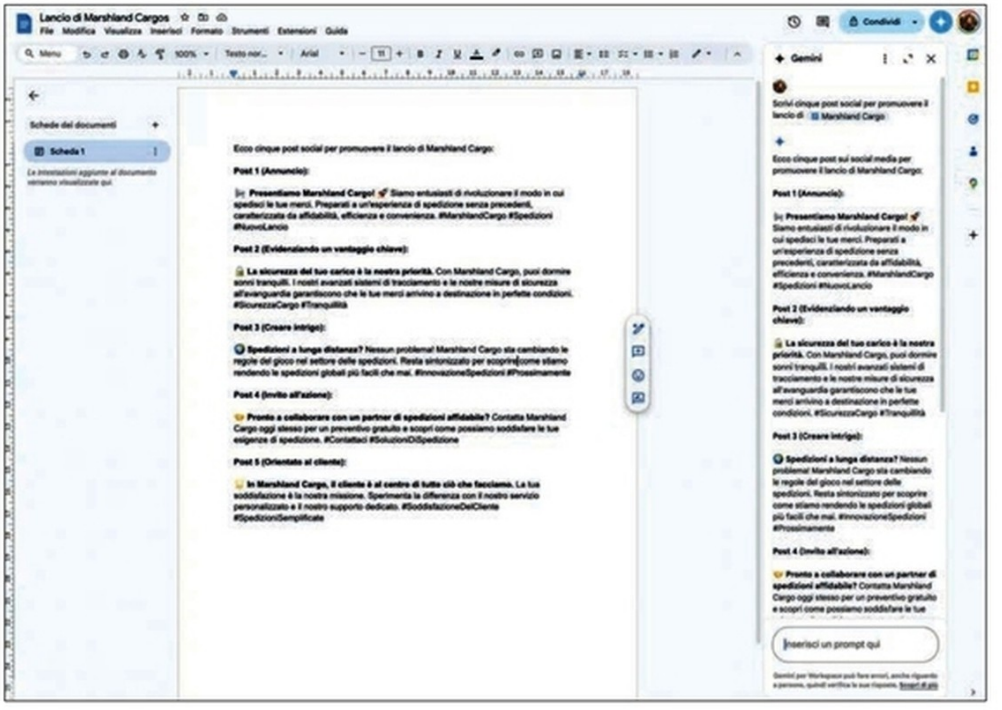
Une intégration au cœur des outils collaboratifs

Avec Gemini, les utilisateurs de Google Workspace peuvent bénéficier de l'IA directement depuis Gmail, Google Docs, Sheets et Drive. L'IA permet, entre autres, de rédiger des e-mails plus rapidement, de générer des rapports complexes ou encore de proposer des analyses de données avancées dans Sheets. Michael Benisty, head of data & AI chez Ledger explique son approche de Gemini pour Workspace : *« Avec Gemini pour Google Workspace, l'IA générative devient accessible en un clic à tous les collaborateurs de Ledger. Intégré à Google Workspace, dans un environnement parfaitement sécurisé, elle permet un gain de productivité, contribue à l'innovation et facilite l'adoption du réflexe GenAI à travers l'organisation. »* Les fonctionnalités de Gemini visent à accélérer les processus de travail tout en améliorant la qualité des résultats obtenus. Pour Frédéric Arnoux, Gemini facilite et transforme l'expérience utilisateur au quotidien : *« Gemini dans Workspace fournit des fonctionnalités avancées et une assistance complète, que cela soit pour trouver des informations plus rapidement, analyser des données complexes pour prendre de meilleures décisions, ou encore créer du contenu de manière efficace. L'IA peut générer par exemple une fiche de poste ou un document de lancement de produit à partir d'idées de départ, en réduisant considérablement le temps nécessaire à ces tâches. Elle offre également un accès élargi à des outils spécialisés pour des*

Frédéric Arnoux, chief technical officer Europe Middle East & Africa de Google Workspace

> *« Le service leur permet de réaliser des documents mieux présentés, mieux organisés et avec plus d'impact. Il simplifie également leurs tâches répétitives et fastidieuses et leur laisse, de fait, plus de temps pour des tâches à forte valeur ajoutée »*

tâches complexes comme la recherche de marché : un chef de produit peut solliciter une analyse comparative pour un lancement, comme celui d'une nouvelle chaussure de sport. Cela inclut des informations sur la taille du marché, les concurrents et les stratégies de prix. La création de contenu devient quant à elle plus fluide en permettant d'éviter le fameux syndrome de la page blanche grâce à des idées préstructurées. Une production enrichie peut ainsi être rapidement développée à partir d'un squelette de document bien défini. »



Une vue de Gemini dans Workspace

Les « Gems » au service de l'automatisation des tâches

Gemini introduit une innovation majeure dans la suite bureautique avec les « Gems ». Ces agents intelligents automatisent des tâches spécifiques pour améliorer la productivité et l'efficacité des équipes. *« Intégrés à Gemini, les Gems sont des modules intelligents qui prennent en charge des tâches spécifiques grâce à l'intelligence artificielle. Ils sont conçus pour répondre aux besoins variés des entreprises, avec deux approches principales : les Gems prédéterminés et les Gems personnalisables. Prêts à l'emploi, les premiers sont conçus par Google pour automatiser des tâches comme la rédaction d'emails à partir d'informations fournies, ou encore du coaching ou du soutien pour des processus métiers spécifiques comme la planification ou la gestion de flux de travail. Avec les seconds, les utilisateurs peuvent créer leurs propres Gems pour répondre à des besoins spécifiques. Un responsable financier peut par exemple concevoir un Gem dédié à l'analyse des budgets mensuels de son entreprise, puis le partager avec son équipe ou dans une bibliothèque pour le rendre accessible à tous. Ces agents simplifient la création et l'adaptation de contenus tout en optimisant leur qualité. »*

Des textes « toujours » originaux

En plus des modules intelligents, Gemini se distingue par sa capacité à générer des contenus entièrement originaux, afin de pallier les préoccupations liées au copyright. Selon Google, Gemini génère systématiquement des textes originaux. Pour chaque requête, les résultats varient, et plus les consignes sont précises, plus les variations sont contrôlées. *« Les utilisateurs raffinent généralement le contenu généré par l'IA pour en faire une version finale en renforçant leur contrôle sur le résultat. En cas de problème lié au copyright, Google s'engage à prendre en charge la défense »*, précise toutefois Frédéric Arnoux. Gemini dans Workspace permet également de générer des images directement dans Gmail, Docs ou d'autres outils de Google. *« Un commercial peut analyser un appel d'offres et générer une présentation visuelle pour y répondre. Les images enrichissent les documents ou supports marketing de manière rapide et pertinente »*, ajoute l'expert.

La guerre des IA dans les suites bureautiques bat son plein

L'intégration de Gemini dans Google Workspace intervient dans un contexte où la compétition entre les grandes entreprises technologiques pour dominer le marché de l'IA générative ne cesse de s'intensifier. Microsoft, avec son « Copilot AI » intégré à Microsoft 365, et d'autres acteurs comme OpenAI ou IBM cherchent également à conquérir ce marché en pleine expansion. Google mise sur la flexibilité et l'adaptabilité de Gemini pour se différencier. En rendant cette technologie accessible en plusieurs langues, l'entreprise convoite un public mondial et diversifié. Cette stratégie linguistique marque une volonté claire d'intégration locale qui se révèle indispensable pour accroître l'adoption de ces solutions d'IA. Une bonne nouvelle pour les utilisateurs qui pourront profiter de solutions de plus en plus efficaces et adaptées à leur culture. □

J.C

Infrastructure

Aiven gère le soubassement des données

D'origine finlandaise, Aiven propose une plateforme unifiée pour les entreprises qui permet de gérer le streaming, le stockage et l'analyse de données sur tous les principaux clouds, le tout en open source.

La promesse d'Aiven est de fournir une plateforme centralisée aux entreprises pour un accès simplifié, ouvert, plus efficient et économique à leurs données, quels que soient leurs mode et lieu de stockage. Les technologies open source jouent un rôle clé dans les offres de services d'Aiven, permettant ainsi la flexibilité et la personnalisation des solutions. Fondés sur le principe de simplification de la gestion des données, les outils d'Aiven sont conçus pour aider les entreprises à tirer parti des technologies open source dans le cloud sans la complexité de la gestion et du maintien des infrastructures. Malgré une assez faible présence en France (une dizaine de personnes), le marché national est stratégique pour l'éditeur. C'est le deuxième marché par revenu après les USA pour l'entreprise.

Toute une gamme de produits pour gérer les données

En s'appuyant sur des briques open source, l'éditeur propose de nombreux produits et services de données autour du streaming, du stockage et de la gestion de l'analyse des données. Ainsi, pour le streaming, Aiven a dans son portefeuille Aiven for Apache Kafka, Aiven for Apache Kafka Connect, Aiven for Apache Kafka MirrorMaker2, Aiven for Apache FLink, Karapace, Klaw. Pour le stockage, Aiven for PostgreSQL, Aiven for MySQL, Aiven for Valkey et Aiven for Dragonfly sont disponibles.

La carte du multicloud

Dimitri Casvigny, vice-président en charge du développement de l'entreprise et du développement durable chez AIven explique : *« La plateforme est présente sur différents clouds et permet de prendre un produit, une base de données ou un protocole de flux de données, comme Apache Kafka et de le positionner dans n'importe lequel de ces clouds »*. Il ajoute : *« AlloyDB Omni de Google, qui est destiné à être déployé dans différents environnements, est sur notre plateforme. C'est très important pour nous parce que cela valide la performance de notre plateforme et le produit peut être déployé chez tous nos partenaires Cloud »*. Aiven est partenaire des hyperscalers américains mais aussi d'OVH et d'autres acteurs locaux, afin de permettre la géolocalisation des données. Il continue : *« Si Google nous a choisis, cela a été réfléchi de notre côté. C'est un produit qui est compatible avec PostgreSQL que nous proposons par ailleurs »*.

Dimitri Casvigny, vice-président en charge du développement de l'entreprise et du développement durable chez Aiven

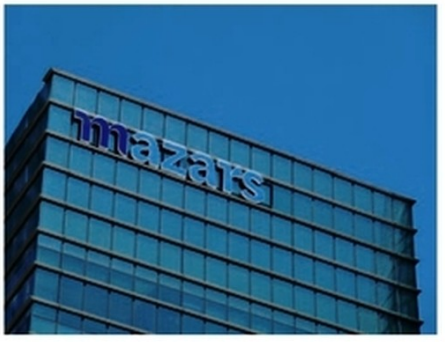
« On a construit notre propre engin d'orchestration qui se base sur les machines virtuelles »

Une architecture virtualisée

« On a fait un choix particulier et on reste sur ce choix. On l'a fait aussi à une certaine époque, quand Kubernetes n'était pas encore mature. On a construit notre propre engin d'orchestration qui se base sur les machines virtuelles. On se rend compte que l'orchestration des containers est fantastique pour tout ce qui est architecture d'application. L'architecture applicative est sur la base de containers. Nos applications sont sur la base de containers.
Mais quand vous voulez orchestrer les récipients de données, on a toujours pensé qu'il y avait un argument très fort pour orchestrer les machines virtuelles. Parce que les machines virtuelles sont persistantes dans le temps. Nous, on orchestre les machines virtuelles. Ça nous donne un certain avantage concurrentiel, assez apprécié de la part des développeurs. Quand vous faites la maintenance ou l'upgrade, généralement, il y a un downtime. Avec nous, il y a zéro downtime. C'est l'une de nos forces », assure Dimitri Casvigny.

La migration dans le Cloud comme point d'entrée

Notre interlocuteur nous aussi précisé que les clients venaient vers Aiven quand il s'agissait de migrer d'un cloud à un autre ou d'une rationalisation de la présence dans le Cloud après une fusion par exemple. C'est dans ces cas que les clients voient la pertinence de la plateforme, afin de limiter les coûts et de gagner en flexibilité et en rapidité dans leur mouvement de données dans le Cloud. ☐ **B.G**



IA Forvis Mazars forme 5 000 salariés aux usages de l'IA

D'ici le printemps 2026, tous les collaborateurs en France du cabinet d'audit et de conseil Forvis Mazars, qui a récemment lancé sa suite d'intelligence artificielle, seront formés aux pratiques de l'IA spécifiques à leur métier.

Le groupe de services d'audit, de conseil et de fiscalité Forvis Mazars forme ses 5 000 collaborateurs en France aux usages de l'intelligence artificielle pour qu'ils l'injectent dans leurs tâches du quotidien. Lancée en octobre 2024, en même temps que son programme d'IA, la formation démarre par les 300 associés. L'objectif est de former tout le monde d'ici la fin du 1er trimestre 2026. *« L'idée est d'adapter nos manières de travailler à un contexte de mutation profonde de nos métiers et ce à l'échelle de toute l'organisation,* met en avant Guillaume Ravix, directeur du développement du capital humain chez Forvis Mazars. *La formation s'articule autour de trois phases : l'acculturation, le passage à l'action et l'approfondissement. »*

Plus de 30 responsables d'unités managériales sont mobilisés afin de spécifier les contenus et les jalons de formation dédiés à chaque activité, y compris les fonctions transverses comme le marketing, les ressources humaines ou la communication. Le groupe souhaite accompagner la transformation des pratiques des salariés et qu'ils gagnent en rapidité.

Démystifier l'IA

Pour Guillaume Ravix, *« l'objectif majeur est que chacun se sente libre de recourir à l'IA, dans un environnement 100 % sécurisé et fiable. La meilleure manière de démystifier l'IA, c'est de la pratiquer au quotidien et ensemble. Nos formations permettent d'embarquer tout le monde dans l'organisation, et que chacun s'en saisisse et la perçoive comme un atout pour sa carrière et son employabilité. Nous valorisons le capital humain : investir sur le développement des compétences constitue sur le temps long un véritable avantage compétitif ».*

LANCEMENT DE L'ACADÉMIE DIGITALE

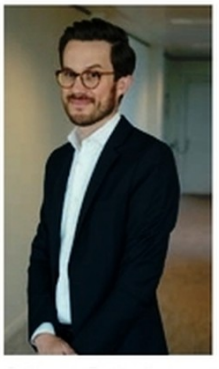
L'Académie digitale a été lancée par Forvis Mazars en décembre 2024 pour diffuser la culture data à tous les niveaux de son organisation. Outre le parcours sur l'IA, elle inclut des formations à la data et au Web3. Elle vise à fournir aux collaborateurs les compétences nécessaires pour évoluer dans leurs métiers tout en s'adaptant aux nouvelles technologies.

LES 3 AXES DE LA SUITE IA

Forvis Mazars a intégré l'intelligence artificielle dans toutes ses méthodes de travail et strates organisationnelles en lançant, en octobre 2024, son programme d'IA baptisé la Suite IA. Il comporte trois axes : l'utilisation de Copilot, l'outil d'IA intégré à Microsoft Office, l'exploitation de l'IA générative développée en interne, dénommée MAIA, pour optimiser les métiers, et son intégration dans les solutions proposées aux clients. L'usage de MAIA permet par exemple d'automatiser et d'optimiser l'analyse de rapports d'audits ou financiers, de proposer des recommandations ou de détecter les anomalies, libérant les collaborateurs de tâches répétitives. Dans le conseil, l'IA sert à traiter efficacement des documents complexes, comme des propositions commerciales ou des synthèses de marché.

Diverses méthodes de formation

La formation inclut des modules d'apprentissage variés : classes virtuelles, modules d'autoformation en ligne, vidéos pédagogiques, mentorat inversé, outils intégrés aux environnements de travail pour des exercices pratiques en temps réel. *« Ces contenus doivent rester accessibles et désirables, et que chacun puisse s'en saisir quel que soit son métier et sa connaissance des outils technologiques »*, ajoute Guillaume Ravix.



Guillaume Ravix, directeur du développement du capital humain chez Forvis Mazars

Mathilde Le Coz, directrice des ressources humaines de Forvis Mazars en France, conclut : *« Dans un contexte de transformation durable de nos métiers, nous avons la responsabilité d'accompagner nos collaborateurs dans l'évolution des compétences, et leur permettons de se former en communauté, pour se concentrer sur les tâches à plus forte valeur ajoutée. »* ☐

C.C

Santé

L'IA veut soulager l'hôpital

Le CHU de Montpellier multiplie les expérimentations pour optimiser la pratique médicale avec des outils numériques. Basé sur des LLM, un assistant virtuel automatise partiellement de nombreuses tâches à partir des notes médicales prises pendant un passage aux urgences.

Pour David Morquin, médecin spécialisé en maladies infectieuses et tropicales au CHU de Montpellier, le recours au numérique dans les pratiques médicales est devenu une évidence dès 2012. À l'époque, l'établissement commençait à utiliser le dossier patient informatisé DPI. Il crée au cours des années suivantes une structure baptisée Erios, pour Espace de Recherche et d'Intégration des Outils numériques en Santé. Ce centre est dédié aux usages du numérique au sein du CHU, avec le but de rapprocher les médecins et la direction. L'objectif est d'exploiter les données médicales pour améliorer les pratiques, notamment à travers le recours au DPI. Les développements attendus portent sur la recherche clinique, le pilotage des activités, la médecine prédictive et la production de documents. Dans ce contexte, David Morquin noue également des liens avec l'université de Montpellier et l'éditeur Dédalus. Le centre Erios obtient un financement en décembre 2022 de 3,3 M€ dans le cadre du programme Santé numérique. Composé de trois collaborateurs de Dédalus, de trois assistants de recherche, d'une ingénieure en sciences du langage et d'un médecin, l'équipe expérimente des cas d'usage visant à alléger les tâches et la charge cognitive des praticiens à partir des informations du DPI (biologie, comptes-rendus…). Seules quelques sources comme l'imagerie médicale demeurent en dehors du dossier patient. Un lien dans le DPI permet alors d'accéder à ces données, de visualiser un scanner par exemple.

En 2023, les premières expérimentations se déclinent sur les cas d'usage comme la supervision des antibiothérapies, des immunosuppresseurs, des antalgiques… *« Toute une série de cas d'usage se décline dans de la data visualisation, en agrégeant les informations et en les hiérarchisant pour faciliter l'identification d'un examen manquant par exemple, ou encore montrer la corrélation entre la prise d'un médicament et l'apparition d'une toxicité sur le foie »*, illustre notre interlocuteur. Les prototypes développés prennent la forme de timeline et de tableaux de bord, et sont par exemple baptisés AntibioVis. Des praticiens du CHU sont impliqués dans les développements pour tester leur praticité, comme le nombre de clics nécessaire. Dans un autre registre, une autre expérimentation vise à automatiser au mieux le codage des séjours des patients, pour faciliter la justification des dépenses et le bilan financier du CHU.

Un autre cas d'usage finalisé aujourd'hui fait l'objet du logiciel IsoPsy qui couvre la prescription et le suivi de l'isolement thérapeutique en psychiatrie. En 2022, en France, cette mesure a concerné 76 000 personnes. *« Les procédures réglementaires à respecter sont très contraignantes pour les équipes. En particulier l'article 17 qui précise les conditions d'hospitalisation sous contrainte et d'isolement thérapeutique »*, souligne David Morquin. Le logiciel a pour objet de favoriser la compréhension par les équipes de ce processus complexe et d'alléger la tâche des soignants en garantissant son respect. *« L'une des difficultés majeures pour les soignants porte sur la transmission d'information entre les équipes ou encore sur l'identification des prochaines étapes »*. L'éditeur Dédalus devrait démarrer l'industrialisation et la commercialisation du logiciel cette année.

IA générative pour oublier le jargon

Outre la data visualisation, l'aide à la planification et au suivi, une autre famille de travaux en cours porte sur l'IA générative, et a été dénommée Erios Assistant. Pour développer ces cas d'usage, l'équipe teste la plupart des LLM, Mistral, Lama3.1, GPT 4… à travers cet « assistant » conversationnel associé à des bibliothèques de

prompts. Les cas d'usage sont variés, et pour l'heure, ne sont pas en production et demeurent des expérimentations supervisées. Point commun, ils prennent tous en compte des notes médicales prises en langage naturel. L'un des pilotes *« prend en entrée ces notes prises lors d'un passage aux urgences pédiatriques et génère une version personnalisée, à destination du jeune patient ou de sa famille expurgée du jargon médical et, traduite si besoin dans la langue du patient. Cette version est relue et doit être validée par un praticien dans l'outil »*, précise David Morquin. D'autres cas ont pour but d'alléger les tâches administratives. Il s'agit alors d'automatiser le remplissage de documents comme les certificats MDPH (handicap) et pour la gériatrie. Un pilote a également pris en entrée les données patients pour identifier ceux susceptibles de correspondre aux critères d'inclusion dans des protocoles médicaux. L'équipe prévoit également de pouvoir identifier plus facilement les patients les plus à mêmes de suivre des actions de prévention, comme l'arrêt du tabac par exemple. Enfin, dans le cadre de l'analyse des retours patients (questionnaire à l'issue d'un séjour), traiter les quelque 27 000 verbatims reçus par an et les classer dans les 21 catégories (fluidité des parcours, accueil, admission, gestion administrative…) constitue un défi et mobilise énormément de ressources humaines. Erios Assistant prend en charge cette étape. *« Nous avons travaillé à fiabiliser le traitement automatisé du langage pour réduire les hallucinations de l'IA, augmenter le nombre de catégories identifiés. Nos résultats montrent que ce classement peut être réalisé par une IA avec autant de pertinence qu'avec un humain »*, se félicite David Morquin.

David Morquin, professeur au CHU de Montpellier et responsable de l'équipe Erios

Sous surveillance

La mise en production de la plupart de ces pilotes pose question notamment en termes d'hallucinations. *« Chaque cas d'usage est examiné par le comité scientifique et éthique »*, insiste David Morquin. Le DPO et le DSI de l'établissement ont également leur mot à dire. Ce d'autant plus que ces IA, surtout les LLM, nécessitent des infrastructures lourdes. Un partenariat avec Dell fournit huit GPU H100, qui servent majoritairement à faire tourner Erios Assistant. D'autres questions sont plus délicates à prendre en compte, en particulier l'impact sur les métiers, le risque de contournement des outils, ou encore les biais liés à l'usage. *« Les interactions entre humains et IA sont un champ d'exploration abyssal. Nous cherchons à en analyser les conséquences potentielles dans la pratique médicale. Par exemple, si un diagnostic est influencé parce que l'IA parle avant le médecin ou non »*, illustre David Morquin. Les impacts sur la relation de soins sont également analysés, ainsi que les conversions pour améliorer l'assistant via du prompt engineering. Enfin, les outils sont ou devraient être optimisés pour réduire l'empreinte carbone liée à chaque requête. Fort de ces premiers cas d'usage, David Morquin a pris la casquette de directeur de la stratégie pour l'IA et la gouvernance des données au sein du CHU en 2024. ☐

Pbr

IA ET AUTOMATISATION : OÙ PLACER LE CURSEUR ?

Dopées par une succession d'appels à projets publics, le dernier en date, *« Industrialisation et Capacités Santé 2030 »*, lancé par BPIFrance court jusqu'en mars 2025, et par l'IA générative, les applications de digitalisation dans le domaine de la santé se sont récemment multipliées. Elles sont notamment censées préfigurer l'hôpital de demain. Des établissements fonctionnant à partir de nombreux processus automatisés pour plus d'efficacité et aussi pallier les "manques" du cerveau humain, en d'autres mots des prothèses numériques. Sur le terrain, l'intérêt de ces outils dépend de chaque cas d'usage. Si une IA bien entraînée reconnaît mieux et plus vite qu'un spécialiste des tumeurs sur de l'imagerie médicale, le bénéfice est loin d'être prouvé pour des tâches nettement plus délicates comme un diagnostic initial. Modéliser tous les cas potentiels reste illusoire. Et avec une approche basée sur l'IA, les hallucinations et le manque d'explicabilité sont des risques bien réels qui peuvent se concrétiser par autant d'erreurs médicales. Bien sûr, le risque reste limité si ces outils se cantonnent à un rôle d'aide à la décision. La question cruciale sera de créer des interfaces homme/IA fluides et de savoir où placer le curseur pour la prise de décision avec ces « boites noires ».

Hyperarme

Anatomie d'une intranquillité planétaire

Auteur reconnu sur le sujet de l'intelligence artificielle, entrepreneur, Flavien Chervet vient de livrer son troisième ouvrage sur le sujet. Après Hyperprompt, Hypercréation, voici donc Hyperarme. Le livre fait à la fois le point sur la technologie de l'intelligence artificielle avec ses bénéfices et ses limites. Il analyse en profondeur les enjeux militaires et géostratégiques de la technologie, ainsi que l'arrivée des systèmes « sur-intelligents ». Il constate ainsi que notre société est à la croisée des chemins avec une alternative : prendre le chemin vers ce qui serait un nouveau « projet Manhattan », en faisant de cette super-intelligence un sujet de compétition mondiale, ou la voie de la coopération internationale pour résoudre les grands problèmes de notre époque. Ce choix n'a pas encore été fait, il est donc encore temps d'agir et, avec l'auteur, de définir un futur qui évite l'auto-réalisation du pire.

NDLR (Flavien Chervet est aussi le frère du PDG du Groupe Ficade, Gaël Chervet, auquel appartient l'Informaticien).

Vers la superintelligence

Le 5 juin 2024, la société DeepMind publie un article intitulé Position: Levels of AGI for Operationalizing Progress on the Path to AGI[1]. Ils y proposent une synthèse du débat et une définition par niveau de performance et de généralité (Fig. 07 en page suivante), menant de systèmes « restreints non intelligents » (pas de performance intelligente et pas de généralité, une calculatrice, par exemple) aux « AGI* compétentes » (performances équivalentes à celles d'un humain moyen dans un grand nombre de tâches cognitives), puis aux… « superintelligences artificielles », abrégées en « ASI » pour « artificial superintelligence » (performances surhumaines dans de nombreux domaines, y compris métacognitifs, et capacités à apprendre de nouvelles compétences).

Nous avons déjà des IA restreintes surhumaines, capables de nous dépasser dans un domaine particulier. AlphaZero en fait partie. La révolution des LLM, tirée par GPT, nous a fait entrer dans la danse de la généralité avec des IA capables de réaliser de nombreuses tâches très diversifiées. Ces IA nous ont fait atteindre le stade « AGI* émergente ». Si les scaling laws se maintiennent, les prochains passages à l'échelle pourraient nous faire atteindre le stade « AGI* compétente », voire « AGI* experte ».

Il est probable que le passage de ces types d'AGI* à l'ASI* ne pose pas vraiment plus de difficultés que le passage d'AGI* émergentes à des AGI* compétentes et expertes.

L'arrivée d'AGI* émergentes comme ChatGPT a déclenché une prise de conscience mondiale et a fait affluer capitaux et talents au service du domaine de l'IA. Si les modèles continuent de s'améliorer entre 2024 et 2028, cet effet sera maintenu, et il faut s'attendre à ce que la bulle s'entretienne. Et si cela nous mène à des AGI* compétentes voire expertes, le monde entier en sera bousculé, et les enjeux seront tels que l'afflux de capitaux et d'intellects sera plus intense que jamais. Le monde entier travaillera à faire advenir l'ASI*.

Et, peut-être, ne serons-nous pas seuls dans cette tâche.

En 1965, alors que l'intelligence artificielle pointe tout juste le bout de son nez[2], le mathématicien I.J. Good a ces mots un brin terrifiants :

« Définissons une machine ultra-intelligente comme une machine capable de surpasser de loin toutes les activités intellectuelles d'un homme, aussi intelligent soit-il. La conception de machines étant l'une de ces activités intellectuelles, une machine ultra-intelligente pourrait

1 : Meredith Ringel Morris et al., Levels of AGI for Operationalizing Progress on the Path to AGI, ArXiv (4 novembre 2023)

2 : Le terme « intelligence artificielle » a été créé en 1956 aux conférences de Dartmouth, pour désigner une branche de l'informatique s'intéressant à la simulation des fonctions cognitives biologiques, notamment humaines.

	Restreinte (capable de réaliser une tâche ou ensemble de tâches clairement délimitées)	**Générale** (capable de réaliser un large spectre de tâches non-physiques, incluant des tâches métacognitives comme l'apprentissage de nouvelles compétences)
Niveau 0 : Non-IA	Non-IA restreinte Exemples : une calculatrice, un compilateur…	Non-IA générale Exemple : automatisation réalisée par des humains, comme Amazon Mechanical Turk…
Niveau 1 : Émergente (Équivalente ou un peu meilleure qu'un humain non qualifié)	IA restreinte émergente Exemples : GOFAI, un système simple à base de règles, comme SHRDLU…	AGI émergente Exemples : ChatGPT, Bard ; Llama 2, Gemini…
Niveau 2 : Compétente (Au moins 50ème percentile des adultes qualifiés)	IA restreinte compétente Exemples : détecteur de toxicité comme JIGSAW, enceinte intelligente comme Siri (Apple), systèmes VQA comme PaLI ; Watson (IBM), LLMs pour un sous-ensemble de tâches (écriture d'essais courts, codage simple)…	AGI compétente Non encore atteinte
Niveau 3 : Experte (Au moins 90ème percentile des adultes qualifiés)	IA restreinte experte Exemples : vérificateur de grammaire et de prononciation comme Grammarly, modèle de génération d'images comme Imagen ou DALL-E 2…	AGI experte Non encore atteinte
Niveau 4 : Virtuose (Au moins 99ème percentile des adultes qualifiés)	IA restreinte virtuose Exemples : Deep Blue, AlphaGo…	AGI virtuose Non encore atteinte
Niveau 5 : Surhumaine (Dépasse 100 % des humains)	IA restreinte surhumaine Exemples : AlphaFold, AlphaZero, StockFish…	Superintelligence artificielle (ASI) Non encore atteinte

Figure 07 : Le tableau proposé par DeepMind dans leur article *Position: Levels of AGI for Operationalizing Progress on the Path to AGI* (Cf. Ress. 23)

concevoir des machines encore meilleures ; il y aurait alors incontestablement une « explosion de l'intelligence », *et l'intelligence de l'homme serait laissée loin derrière. Ainsi, la première machine ultra-intelligente est la dernière invention que l'homme devra jamais faire, à condition que la machine soit assez docile pour nous dire comment la garder sous contrôle. »*[3]

Cette idée d'« explosion de l'intelligence » fait date, et devient rapidement un classique de la science-fiction[4]. Jusqu'à récemment, il s'agissait seulement d'imaginaires fascinants propices aux histoires les plus rocambolesques.

Mais l'exemple d'AlphaDev rappelle que des précédents existent d'IA capables de découvrir de nouvelles techniques informatiques[5]. Et avec *The AI Scientist*, des systèmes informatiques deviennent capables de faire de la recherche scientifique en autonomie. La maturité des LLM* utilisés étant celle de 2024, aucune idée n'était à ce stade une percée, mais qu'inventera une telle architecture logicielle combinée avec une AGI* experte, capable de raisonner de manière rigoureuse, de mobiliser la connaissance de l'ensemble des savoirs humains en science de l'IA, et d'utiliser des processus créatifs plus féconds pour les dépasser, comme ceux développés par DeepMind ?

En outre, les recherches en théorie de l'innovation sont nombreuses depuis une cinquantaine d'années, et le processus de production de nouvelles idées est de mieux en mieux compris et modélisé[6]. Cette *« science de la découverte »* est encore méconnue du grand public, et persiste dans l'imaginaire collectif l'idée d'une « magie » de l'intuition créative.

La créativité artificielle avance pourtant bon train et produit déjà des résultats dans tous les domaines de la science. Automatiser la science et l'exploration de l'inconnu commence à être une réalité.

Enfin, ce scénario possède une puissance inédite. Il réalise deux mythes profonds — peut-être les plus profonds — qui habitent la civilisation occidentale : donner la vie[7] et faire advenir Dieu sur Terre.

On trouve des traces du premier depuis la mythologie grecque : dans sa forge, Héphaïstos a conçu des créatures de roche et de fer qui travaillent pour lui. Elles sont capables d'une grande autonomie pour gérer la forge à sa place lorsqu'il s'absente. On retrouve cette idée dans le mythe du golem de la kabbale juive, dans la légende de Frankenstein et, bien sûr, dans la figure du robot moderne. L'intelligence artificielle est un projet qui dispose

3 : I.J. Good, Speculations Concerning the First Ultraintelligent Machine, Advances in Computers, vol. 6 (1965)

4 : Un autre terme souvent utilisé dans la littérature d'anticipation est « take-off » ou « décollage ».

5 : Un autre exemple est AlphaTensor, aussi conçue par DeepMind, qui en 2022 découvre des algorithmes de multiplica tion de matrices, une opération mathématique massivement utilisée en IA, plus efficaces que ceux qui existent alors.

6 : Epoch AI eux-mêmes ont publié, en mai 2024, un article technique mais très complet sur la modélisation mathématique de l'innovation et sur l'utilisation de différents systèmes d'IA en découverte scientifique.

7 : Un mythe bien masculin, convenons-en… Il paraît que les femmes en ont moins besoin.

de l'énergie émotionnelle d'un enfantement à l'échelle de l'humanité.

Le second est associé à l'idée fabuleuse de transcendance, d'un *« quelque chose qui nous dépasse »*. La transcendance est l'idée centrale qui justifie les religions. L'idée d'explosion de l'intelligence nous rend accessible ce mythe. Elle lui apporte une solution. Une IA s'améliorant toute seule : voilà un mécanisme pour faire advenir la transcendance du divin sur Terre. Tout à coup, l'être humain peut toucher du doigt l'infini.

À elle seule, cette double dimension mythique de la superintelligence* transforme la quête capitaliste d'AGI* capables en une quête spirituelle d'ASI* sacrées.

L'auteur et philosophe controversé Nick Land a forgé le terme « hyperstition ». Celui-ci décrit la force transformative que certaines idées — notamment les mythes — ont sur le cours historique des sociétés. Les scaling laws en général, et l'explosion de l'intelligence en particulier, sont des exemples typiques d'hyperstition. Parce qu'elles raisonnent avec un fantasme mythique profond, ces idées sont performatives. Elles nous font entrevoir l'avenir selon leurs couleurs et nous font agir pour les réaliser. À travers nous, elles se font advenir elles-mêmes. La dimension fantasmagorique de la superintelligence* n'est pas à négliger : elle fournit une énergie incroyable à l'humanité, tendue en direction de sa réalisation. L'envergure et la force des religions prouvent que la transcendance nous rend capables de faire advenir les plus grands projets, en dehors de toute raison.

Le futurologue Raymond Kurzweil s'est approprié ce mythe sous le nom de « singularité » : il s'agit du point dans le temps à partir duquel, la technologie se développant par elle-même en autonomie, les progrès sont si fulgurants qu'il devient parfaitement impossible de prédire ce qu'il adviendra. Depuis maintenant 20 ans, Kurzweil prédit que la singularité arrivera en 2045. Ces dernières années, il s'est mis à envisager de revoir ses prédictions… à la baisse ! Le récent mouvement techno-utopique de l'accélérationnisme efficace[8], quant à lui, appelle de ses vœux une maximisation du progrès technologique et de la dépense énergétique associée pour atteindre au plus vite cette même singularité. Ce mouvement est tout sauf marginal. Il prend de l'ampleur aux États-Unis, connaît dans ses rangs des investisseurs de renom comme Peter Thiel (cofondateur de Paypal et de… Palantir !), et possède des ramifications jusque dans la Maison-Blanche. Les chercheurs qui travaillent sur l'intelligence artificielle eux-mêmes sont en grande partie fascinés par le mythe de la singularité.

Aschenbrenner reprend à son compte ce scénario puissant et renouvelle l'argumentaire à la lumière des dernières avancées en IA. Dans la même veine que Good, il s'appuie sur l'idée qu'un système d'IA capable d'automatiser les tâches cognitives d'un être humain pourrait, en particulier, automatiser le travail d'un chercheur en IA.

Il pousse même le raisonnement plus loin, en s'appuyant sur les caractéristiques de prolifération de l'IA que nous avons déjà évoquées. La puissance de calcul faramineuse nécessaire pour obtenir une AGI* en 2027/2028 est celle nécessaire à son entraînement. Mais, une fois celle-ci entraînée, son utilisation est comparativement très peu énergivore. En outre, elle peut très facilement être dupliquée (il suffit de lancer plusieurs fois le programme, comme on peut lancer plusieurs fois son navigateur ou le logiciel Word). Il faut donc s'attendre à avoir non pas une, mais des millions d'IA « chercheuses en IA », qui travaillent en parallèle sur toutes les pistes de recherche disponibles, et qui soient à l'origine d'innovations permettant de construire des systèmes d'IA plus performants… eux-mêmes capables de meilleures recherches en IA.

Au-delà de la recherche en IA elle-même, l'utilisation de systèmes d'IA avancés devrait produire une accélération de tous les autres domaines de la science. La convergence avec la biologie qui, comme nous l'avons vu, se révèle être majoritairement affaire d'information et de calcul, devrait s'intensifier. Ce devrait aussi être le cas pour la physique elle-même, dont la face numérique se développe considérablement. Toutes ces avancées devraient rétroagir pour accélérer en retour la recherche en IA elle-même, par exemple en permettant l'arrivée de meilleurs processeurs et de centrales énergétiques plus performantes. Ce mécanisme de rétroaction positive est appelé *« loi des retours accélérés »* par R. Kurzweil, qui y voit le contrepoids des frictions imposées par le réel et postulées par ceux qui doutent du maintien des lois d'échelle.

Le domaine de la robotique montre de manière frappante cet effet rétroactif qui produit un moteur de croissance accélérée.

La robotique a longtemps été considérée comme le goulot d'étranglement de l'IA dans le monde physique. Or, elle connaît, depuis le boom des IA génératives, une accélération rapide. Un coup d'œil aux vidéos du robot Figure, dévoilées en 2024, suffisent pour s'en convaincre. Et pour cause, une grande partie des problèmes auxquels est confronté le domaine de la robotique relèvent en fait de questions traitables par l'intelligence artificielle[9].

Le développement de la robotique grand public, en multipliant le nombre de robots dans des environnements variés, permettra de collecter massivement un nouveau flux de données précieux que l'on ne trouve pas sur Internet : des données d'interaction dans le monde physique. Ces données seront clés pour développer des

8 : Qui, au demeurant, se réclame de l'héritage intellectuel de Nick Land. La boucle est bouclée.

9 : Voilà un exemple. Les voitures autonomes étaient jusqu'à présent équipées de LIDAR, des systèmes de détection par laser très sophistiqués et coûteux servant à modéliser l'environnement autour de la voiture. L'humain n'a pas besoin de capteurs d'une telle complexité pour conduire. Deux caméras (ses yeux) suffisent, car son intelligence est capable d'interpréter efficacement les données perçues. La vision comme la navigation dans un environnement sont affaire d'intelligence. La tendance dans les voitures autonomes est à la réduction des capteurs et de la complexité du système physique, au profit de meilleurs algorithmes d'IA embarqués.

systèmes d'IA dotés d'une compréhension plus intuitive et fine de notre monde (un « bon sens »). Ces systèmes permettront à leur tour de rendre les robots plus efficaces. Sur son site web, la société Figure est explicite sur cet objectif. Elle se définit non pas comme un constructeur de robots, mais comme un « moteur de données », et présente son modèle selon le schéma suivant : fabriquer des robots pour collecter des données, qui amélioreront les systèmes d'IA, qui amélioreront les robots, qui deviendront alors utiles dans plus de domaines, ce qui justifiera en retour de fabriquer plus de robots, et ainsi de suite[10].

La valeur captée par les entreprises fabriquant des robots n'est pas celle de la vente de l'objet robot, mais celle de la possession des données précieuses collectées par celui-ci. Leur intérêt n'est pas de vendre cher, mais de mettre le plus de robots possible en circulation. Il faut donc s'attendre à ce que les robots grand public soient vendus à des prix très abordables, et à ce que leur diffusion soit rapide dans les prochaines années, accélérant d'autant le domaine de l'IA.

Au regard de ces mécanismes de rétroaction positive, Aschenbrenner considère qu'il faut envisager non seulement que l'amélioration des systèmes d'IA continue, mais même qu'elle accélère. Avec cette hypothèse en main, il continue à compter les ordres de grandeur et poursuit la courbe au-delà de 2028. Il suppose ainsi que des systèmes dits superintelligents* verront le jour d'ici la fin de la décennie[11].

L'argumentaire reste hautement spéculatif. Mais de nombreux signaux semblent converger, et il serait imprudent de balayer ce scénario du revers de la main.

Toutefois, réaliser un bond supplémentaire, passer de l'AGI* à l'ASI*, se heurterait de nouveau à la limite physique de l'énergie.

On peut facilement envisager que les réseaux de neurones deviennent encore plus gros. On peut imaginer que la pompe financière continue et même augmente son débit. On peut imaginer que l'augmentation de la production mondiale de puces et l'amélioration de leur performance suffisent à alimenter les projets d'ASI*. On peut imaginer que le talent des chercheurs, peut-être couplé à l'amélioration récursive des IA, fournisse les innovations algorithmiques permettant de franchir le mur de la donnée et de continuer à faire croître les performances des IA selon les scaling laws*. Mais peut-on vraiment imaginer construire des centres de calcul encore plus grands et gourmands que ceux nécessaires à l'AGI*, qui engloutiront déjà probablement à eux seuls la puissance énergétique d'un État entier ?

Construire un centre de calcul de 1 à 10 GW pour un coût de quelques dizaines de milliards de dollars est ambitieux, mais à la portée des fonds colossaux des grandes entreprises technologiques. Construire un cluster de 100 GW, libérer la douce somme de 1000 milliards de dollars, et, au détriment de toute raison écologique, aspirer plus de 20 % de la production énergétique des ÉtatsUnis pour le faire tourner, est d'un tout autre acabit[12].

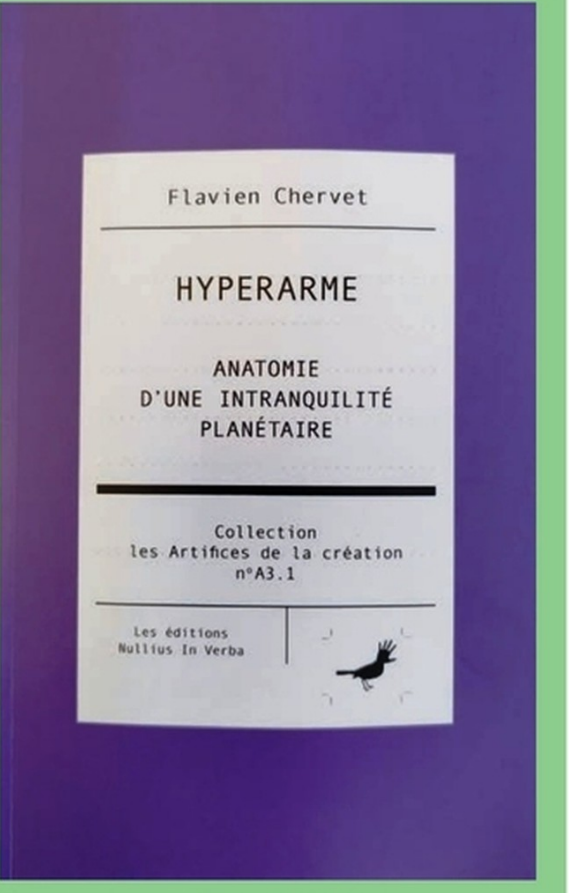
Orienter un effort qui mobilise l'ensemble du tissu industriel des États-Unis pour mettre en œuvre une idée qui reste de l'ordre du théorique : un tel projet est évidemment déraisonnable, mais aussi tout à fait impossible.

Sauf si... □

10 : Source : https://www.figure.ai/ai

11 : Le lecteur intéressé par les détails de l'argumentaire pourra se référer au deuxième chapitre de Situational Awareness intitulé From AGI* to Superintelligence: the Intelligence Explosion.

12 : Leopold Aschenbrenner, Situational Awareness, p. 77 (juin 2024)



Flavien Chervet
Edition Nullius inVerba
ISBN : 979-10-92564-49-5
Prix 19 €

IA 2025 sera-t-elle l'année de l'agentic AI ?

Vous avez aimé l'émergence du machine learning et du deep learning, puis l'explosion de l'IA générative ? Vous allez adorer l'acte 3 : l'arrivée de l'agentic IA. Cette nouvelle approche pourrait bien bousculer le monde du software tel qu'on le connaît.

« *SaaS is Dead !* », c'est comme cela qu'a été relayée, sur les médias sociaux, une interview de Satya Nadella sur le podcast BG2 en décembre dernier... Si le CEO de Microsoft n'a pas été aussi catégorique dans ses propos, il a néanmoins souligné que l'IA et le rôle croissant des agents « intelligents » allaient bouleverser la manière de consommer les applications SaaS. Le schéma base de données/process métiers et frontend des applications Web classiques va être clairement bousculé par l'apparition d'agents logiciels qui vont littéralement réaliser les transactions au nom des utilisateurs.

Cette tendance s'appelle l'agentic AI ou IA agentique. La définition qu'en donne le Gartner fait un peu froid dans le dos : « *L'IA agentique introduira une main-d'œuvre numérique axée sur les objectifs, qui élaborera des plans et prendra des mesures de manière autonome — une extension de la main-d'œuvre qui n'a pas besoin de vacances ou d'autres avantages* ». Les analystes du cabinet américain considèrent que le tiers des applications d'entreprise intégreront l'IA agentique à l'horizon 2028, et que 15 % des décisions professionnelles quotidiennes seront prises de manière autonome... Bien évidemment, de nombreux éditeurs se positionnent déjà sur ces IA agentiques. Les éditeurs d'automatisation des processus et de RPA, mais tous les éditeurs SaaS doivent se positionner aussi, au risque de se faire sortir du marché.

Thomas Husson,
VP Principal Analyst
chez Forrester

« L'IA agentique se développe beaucoup aux États-Unis et va arriver en France en 2025. Ce sera, dans un premier temps, sur la partie logicielle, avec en cascade, un agent qui va effectuer des tâches dans un process. On n'est pas encore sur la partie grand public où on restera encore sur le commerce conversationnel plus classique, avec un chatbot de deuxième génération »

La connaissance métier glisse vers l'IA

Si le rôle de l'humain va être redéfini par l'arrivée de ces agents dopés à l'IA générative, le rôle même des applications est questionné : « « *Nous avons des licences pour toutes ces applications SaaS que nous utilisons à peine... La logique d'entreprise va aller vers des agents qui seront multi-backend et sans discrimination : ils vont mettre à jour de multiples bases de données et toute la logique métier sera dans le niveau d'IA* », estime Satya Nadella. Celui-ci pointe aussi les nombreux défis qu'il faudra relever avant que ce modèle d'IA agentique puisse s'imposer. Il faudra mettre en place de nouveaux business models en remplacement des licences utilisateurs « humains », résoudre des problématiques de cybersécurité évidentes, puisque ces IA capables d'agir à la place d'humains pourront, par exemple, lancer des ordres d'achat ou des virements bancaires automatiquement...

Un autre effet de l'IA agentique est de rendre le remplacement des backends beaucoup plus simple. Si votre interface utilisateur est une IA multi-applications avec laquelle vous interagissez en langage naturel, passer d'un ERP à un autre devient beaucoup plus simple. Or, à peu près toutes les applications d'entreprises sont potentiellement concernées par ce changement d'approche.

Dans le B2C, l'IA agentique a fait son arrivée sous la forme du commerce conversationnel : on discute avec un agent pour trouver le bon produit dans le catalogue du marchand. Demain, on demandera à son agent personnel de parcourir tous les sites marchands pour trouver la paire de baskets correspondant aux sports pratiqués, à ses gouts en termes de forme ou de couleur, aux contraintes de prix, et l'agent effectuera l'achat de manière autonome lorsque tous les critères auront été remplis. □

A.C

Service public

L'eau sous pression

Le pôle de compétitivité Aqua-Valley œuvre pour améliorer la gestion de l'eau. Les entreprises qu'il soutient mettent largement à contribution l'IA et les données issues des satellites pour améliorer le pilotage de la ressource. En décembre dernier, les Aqua businness days ont donné un aperçu des avancées dans le domaine.

En 2023, l'eau potable a été livrée par camions citernes dans plusieurs localités du sud de la France par camions citernes en raison de la sécheresse. Une alerte sur cette ressource qui ne date pas d'hier. *« Dès 2014, la Commission européenne avait demandé aux régions de définir, après benchmark, leurs domaines de spécialité d'innovation pour l'eau »*, rappelle Yvan Kedaj, directeur du pôle Aqua-Valley, un pôle de compétitivité regroupant environ 250 adhérents (entreprises, organismes de recherche et de formation et association) localisés en Occitanie et Provence-Alpes-Côte d'Azur. *« L'Occitanie a été la seule région européenne reconnue de spécialité sur la totalité des problématiques liées à l'eau,* ajoute-t-il. *Ces sujets d'innovation sont travaillés en région avec Éa éco-entreprises (notre délégation en région Provence-Alpes-Côte d'Azur), et au niveau national avec notre homologue, le pôle Aquanova (sur les régions Centre-Val de Loire et Grand-Est), avec lequel nous formons les pôles EAU »*. Le pôle a pour objectif d'aider et de mettre en relation tous les acteurs concernés par la gestion de l'eau, et de faciliter l'identification ou le développement de solutions permettant de s'adapter au changement climatique.

Les projets s'étendent de la métrologie au pilotage, en passant par la réutilisation, le traitement et le développement d'outils prédictifs sur les risques. Ils impliquent classiquement des startups et le monde académique. À l'origine, plutôt basés sur des SIG (système d'information géographique) et sur de la modélisation, ils utilisent aujourd'hui massivement IA et données. Le partenaire académique majeur du pôle, le laboratoire HydroSciences de Montpellier (CNRS et Institut de recherche pour le développement) mène des travaux pour mieux comprendre, et anticiper les impacts du climat et des activités humaines sur les ressources en eau des régions méditerranéennes et tropicales. Ses chercheurs modélisent le cycle de l'eau dans sa totalité, de l'atmosphère à la rivière en passant par la végétation et le sol. Ils travaillent aussi sur les contaminants métalliques et émergents, et les risques sanitaires liés aux bactéries pathogènes hydriques. Les résultats prennent entre autres la forme de logiciels comme le logiciel SW2D. Celui-ci a pour objectif de modéliser l'hydrodynamique des crues et des lacs. Les équipes ont développé une nouvelle version ne nécessitant pas le recours au HPC, mais optimisée pour tourner sur du Nvidia. Les chercheurs proposent en open source d'autres outils comme Athys, développé à l'IRD et destiné à des applications diverses : gestion de la ressource en eau, prévision des événements extrêmes, études d'impacts liés à des modifications anthropiques ou climatiques. Parmi les projets en cours, Starwars travaille sur les réseaux d'eau urbains pour compléter la collecte des données dans ce contexte, et améliorer les connaissances.

ÉVITER LA GUERRE DE L'EAU

Organisées par Aqua-Valley et Éa éco-entreprises, la 4e édition des Aqua Business Days a vu intervenir François Gemenne, membre du GIEC et spécialiste dans la géopolique de l'environnement sur le thème "Comment éviter les conflits autour de l'eau". Une fois le décor dressé, les journées ont laissé une large place aux avancées technologiques pour mieux prédire et piloter cette ressource. Trois projets, Récolt'Ô, Electrotate, menés avec le labo Gepea du CNRS, et VigiNappe pour mieux connaître l'état de la ressource en eau souterraine, ont été récompensés par les Trophées Aqua-Valley.

UNE **MODERNISATION DES RÉSEAUX** COMPLIQUÉE

Sur le terrain, les projets de modernisation des réseaux d'eau potable ou d'assainissement tiennent encore souvent aux seules évolutions réglementaires et à l'abandon du RTC. Une première étape déjà compliquée quand les moyens des collectivités chargées de la gestion de cette ressource demeurent limités. Chargés, par exemple, d'actionner les pompes de relevage en fonction du niveau dans les châteaux d'eau, *« les automates de télégestion dialoguant auparavant via le RTC doivent communiquer via des réseaux numériques. La complexité technique majeure tient à la multiplicité des protocoles à prendre en compte sur ces réseaux, que ce soit la fibre optique, les liaisons filaires privées, les réseaux 4G/5G/LTE-M/NB-IoT/LPWA (Low Power Wide Area) »*, explique Marie-Armelle Bories, dirigeante de Dralam, un cabinet d'étude spécialisé dans l'IoT. Il s'agit également de prendre en compte les exigences réglementaires récentes en termes de cybersécurité via du cryptage au travers de VPN (Virtual Private Network). *« Une complication supplémentaire existe quand nos clients, petites structures ou collectivités, ont des DSI compétents en IT, mais qui ne sont pas familiers des réseaux numériques complexes »*, ajoute la dirigeante.

D'autres projets récents utilisent massivement les données, grâce à la multiplication des IoT et aux données issues des satellites. Le satellite SWOT* (Surface Water and Ocean Topography), lancé en orbite à 890 km fin 2022, apporte une précision inédite dans l'observation des courants océaniques, du niveau de la mer, et aussi des réserves d'eau douce et des principaux cours d'eau. Une précision due à un nouvel instrument, baptisé KaRin, doté de deux antennes situées aux extrémités d'un mât rigide de 10 mètres de long. *« Diverses structures, les scientifiques, mais aussi les agences de l'eau et les départements ont créé et alimentent des bases de données en accès libre. Des innovations sont développées en matière de métrologie des eaux souterraines. Le défi actuel est d'améliorer le prédictif à partir de toutes ces sources avec les apports de l'IA »*, décrit Yvan Kedaj.

Marie-Armelle Bories, dirigeante de Dralam


Yvan Kedaj, directeur du pôle Aqua-Valley

Outre la seule mesure quantitative, il s'agit également de mieux gérer cette ressource. Dans ce but, le pôle facilite l'opérationnalisation des résultats de recherche et le soutien aux entreprises. Illustration, il vient de récompenser au cours des Aqua Business Days (voir encadré), le projet Récolt'Ô porté par Makina Corpus, en partenariat avec le CSTB. Celui-ci a pour objectif d'aider au déploiement de récupérateurs d'eau sur tout type de bâtiment, en déterminant le volume de cuve adéquat en fonction des prévisions de pluviométrie locale et des usages locaux de l'eau. Plus globalement, le projet Reut-O-Sud cherche à développer la filière de la réutilisation des eaux usées traitées — REUT dans la région Sud PACA. *« Des modèles de gouvernance de la REUT sont en train de se mettre en place »*, avance Yvan Kedaj.

Les problématiques liées à l'eau ne se limitent pas à la métrologie et au pilotage. Société spécialisée dans les domaines liés à l'eau, BRL Ingénierie a créé la filiale Predict, dédiée à la prédiction des inondations et autres risques hydrométéorologiques. Les deux tiers des 36 000 communes françaises sont exposées à ces risques naturels. Destiné aux collectivités, Predict Services avertit par tous les moyens (SMS, Internet...) en cas de phénomènes hydrométéorologiques à risque menaçant un territoire. Des alertes basées sur les travaux de recherche menés par Predict avec Météo France et des données issues de satellites. Autre aspect vital, les projets visent également à surveiller la qualité de l'eau potable et à améliorer l'assainissement. *« Les systèmes membranaires sont de plus en plus utilisés, en raison de leur capacité à filtrer jusqu'aux virus, notamment »*, souligne Yvan Kedaj. A la base d'un nouveau processus électrochimique d'oxydation permettant de traiter les eaux usées sans ajouter de produits chimiques, la société Electrotate est soutenue par Aqua-Valley. L'innovation de ce processus repose sur la mise en mouvement des électrodes. De nombreux défis demeurent, techniques et économiques. Les communes, responsables de ces ressources n'ont, pour la plupart, pas les moyens de mettre en place les systèmes de filtration les plus récents. Et dans le registre technique, les approches capables non seulement de filtrer les polluants, mais aussi de décomposer les molécules de synthèse restent largement à découvrir. □

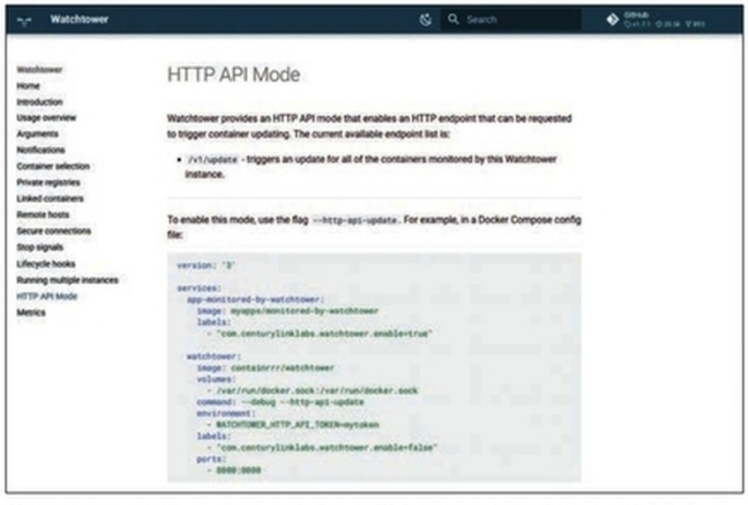
PBr

Automatisation

Mettre à jour automatiquement ses containers Docker avec Watchtower

Une fois vos différents containers installés sur Docker, il reste une partie assez fastidieuse : leur mise à jour. En fonction des projets, vous pouvez avoir des mises à jour hebdomadaires, voire plus fréquentes, à réaliser, ce qui peut vite devenir très chronophage. Dans cet article, nous allons vous présenter un outil permettant d'automatiser ce processus : Watchtower.

L'utilisation de Docker apporte de nombreux avantages, comme notamment la possibilité d'empaqueter des applications dans des unités standardisées pour le développement logiciel. Cela simplifie grandement le déploiement et la mise à l'échelle des applications. Néanmoins, la gestion des mises à jour des images Docker sur un serveur peut devenir pénible. Vous devez récupérer manuellement les nouvelles versions d'une image et redémarrer son conteneur à chaque fois qu'une mise à jour est publiée. C'est là que Watchtower entre en scène. Cette solution, basée elle aussi sur un conteneur, surveille les conteneurs Docker en cours d'exécution et guette les éventuelles modifications apportées aux images qui les composent. Lorsque Watchtower détecte qu'une image a été mise à jour, il arrête automatiquement les conteneurs associés, afin de récupérer la nouvelle image et les redémarre avec les mêmes options que lors du déploiement initial.



Vous pouvez consulter la documentation de Watchtower sur le site du projet, à l'adresse https://containrrr.dev/watchtower/introduction

Installation

Watchtower est distribué sous forme d'image Docker hébergée sur Docker Hub. Son installation est aussi simple que l'exécution d'un conteneur à partir de cette image. Il faut tout d'abord télécharger la dernière image Watchtower avec l'option pull de docker :

```
docker pull containrrr/watchtower
```

Une fois le téléchargement terminé, il s'installe via l'instruction docker run suivante :

```
docker run -d --name watchtower \
-v /var/run/docker.sock:/var/run/docker.sock \
containrrr/watchtower
```

Voyons à quoi correspondent les options transmises à la commande docker run :

- -d : Exécute le conteneur Watchtower en mode détaché
- --name watchtower : Nomme le conteneur "watchtower", ce qui facilitera son identification
- -v /var/run/docker.sock:/var/run/docker.sock : Monte le socket Docker dans le conteneur Watchtower, afin qu'il puisse communiquer avec le daemon Docker. Cela permettra au code de Watchtower d'interagir avec l'API Docker, afin de contrôler les containers en cours d'exécution.
- containrrr/watchtower : C'est l'image Docker Watchtower à utiliser

Cette instruction créera et démarrera un conteneur Watchtower en arrière-plan, qui pourra aussitôt commencer la surveillance des autres conteneurs. Vous pouvez aussi exécuter le docker-compose suivant :

```
version: "3"
services:
  watchtower:
    image: containrrr/watchtower
    volumes:
      - /var/run/docker.sock:/var/run/docker.sock
```

Registres privés

Si les images sont extraites de registres Docker privés, les credentials d'authentification au registre doivent être fournies via les variables d'environnement REPO_USER et

REPO_PASS ou alors en montant le fichier de configuration docker de l'hôte dans le container (précisément à la racine du système de fichier du conteneur, soit /).

Voici comment passer les variables d'environnement :

```
docker run -d \
 --name watchtower \
 -e REPO_USER=username \
 -e REPO_PASS=password \
 -v /var/run/docker.sock:/var/run/docker.sock \
 containrrr/watchtower container_to_watch -debug
```

Notifications des mises à jour de conteneurs

Avec la configuration par défaut, Watchtower vérifie silencieusement les mises à jour des images de conteneur en arrière-plan sans vous en informer. Vous pouvez modifier ce comportement en définissant certaines variables d'environnement lors du lancement de Watchtower. Si vous souhaitez être notifié lorsque Watchtower met à jour les conteneurs, vous devez passer la variable d'environnement -e WATCHTOWER_NOTIFICATIONS=email. D'autres options sont disponibles comme slack, msteams, gotify et autres. slack permet de publier des notifications sur Slack. Vous devrez fournir l'URL du webhook et le canal. msteams sert à envoyer des notifications via Microsoft Teams. Il faut là aussi fournir une URL de webhook. gotify est bien entendu utilisé pour envoyer des notifications via Gotify. Cette option nécessite l'enregistrement d'une app Gotify, ainsi que la spécification d'une URL de serveur et d'un jeton d'application. Cela donnera quelque chose comme cela pour cette dernière option :

```
docker run -d \
 --name watchtower \
 -v /var/run/docker.sock:/var/run/docker.sock \
 -e WATCHTOWER_NOTIFICATIONS=gotify \
 -e WATCHTOWER_NOTIFICATION_GOTIFY_URL="https://my.gotify.tld/" \
 -e WATCHTOWER_NOTIFICATION_GOTIFY_TOKEN="SuperSecretToken" \
 containrrr/watchtower
```

Si vous voulez désactiver la vérification TLS pour l'instance Gotify, vous pouvez le spécifier soit avec la syntaxe -e WATCHTOWER_NOTIFICATION_GOTIFY_TLS_SKIP_VERIFY=true soit avec l'option --notification-gotify-tls-skip-verify.

Notifications au démarrage et à l'arrêt

Vous pouvez demander d'être notifié lorsque le conteneur Watchtower démarre et s'arrête en passant l'option -e WATCHTOWER_NOTIFICATIONS_LEVEL=start-exit, comme ceci :

```
docker run -d --name watchtower -v /var/run/docker.sock:/var/run/docker.sock \
-e WATCHTOWER_NOTIFICATIONS=email \
-e WATCHTOWER_NOTIFICATIONS_LEVEL=start-exit containrrr/watchtower
```

De cette manière vous recevrez des notifications lors des mises à jour de conteneurs mais aussi au démarrage et à l'arrêt de Watchtower.

Configuration des services de notification

L'activation des notifications implique de configurer les services concernés. Cela se fait, soit en passant des variables d'environnement supplémentaires, soit en montant des fichiers de configuration YAML dans le conteneur.

Configuration de l'e-mail

Pour recevoir des notifications par e-mail, vous devez fournir les informations suivantes :
WATCHTOWER_EMAIL_FROM : L'adresse depuis laquelle les e-mails de notification seront envoyés
WATCHTOWER_EMAIL_TO : L'adresse de destination des notifications
WATCHTOWER_EMAIL_SERVER : L'adresse du serveur SMTP
WATCHTOWER_EMAIL_SERVER_PORT : le port du serveur SMTP
WATCHTOWER_EMAIL_SERVER_USER : le nom d'utilisateur SMTP
WATCHTOWER_EMAIL_SERVER_PASSWORD : le mot de passe SMTP

AUTOMATISER LE DÉMARRAGE DE WATCHTOWER

Si Watchtower met à jour automatiquement les nouvelles images Docker, il faut aussi s'assurer qu'il se lancera automatiquement lors du démarrage du daemon Docker. Vous pouvez pour cela définir un simple fichier d'unité systemd. Créez le fichier /etc/systemd/system/watchtower.service en intégrant ces quelques lignes :

```
[Unit]
Description=Watchtower - Mise à jour automatique des conteneurs Docker
Requires=docker.service
After=docker.service
[Service]
Restart=always
ExecStart=/usr/bin/docker start -a watchtower
ExecStop=/usr/bin/docker stop -t 2 watchtower
[Install]
WantedBy=multi-user.target
```

Rechargez ensuite systemd et n'oubliez pas d'activer le service Watchtower pour qu'il se lance au démarrage :

```
sudo systemctl daemon-reload
sudo systemctl enable watchtower
```

Ainsi Watchtower démarrera automatiquement à chaque redémarrage du serveur docker.

Avec une adresse gmail, cela donnerait quelque chose de ce genre :

```
docker run -d --name watchtower -v /var/run/docker.sock:/var/run/docker.sock \
 -e WATCHTOWER_NOTIFICATIONS=email \
 -e WATCHTOWER_EMAIL_FROM=addressexp@gmail.com \
 -e WATCHTOWER_EMAIL_TO=addressdest@undomaine.fr \
 -e WATCHTOWER_EMAIL_SERVER=smtp.gmail.com \
 -e WATCHTOWER_EMAIL_SERVER_PORT=587 \
 -e WATCHTOWER_EMAIL_SERVER_USER=addressexp@gmail.com \
 -e WATCHTOWER_EMAIL_SERVER_PASSWORD=SMTP_password \
 containrrr/watchtower
```

Configuration de MS Teams

C'est plus simple pour les notifications MS Teams. Vous n'avez à fournir qu'une seule information, votre URL de webhook MS Teams :

```
WATCHTOWER_NOTIFICATIONS_MSTEAMS_WEBHOOK_URL -
Cela donnera, par exemple :
docker run -d --name watchtower -v /var/run/docker.sock:/var/run/docker.sock \
 -e WATCHTOWER_NOTIFICATIONS=msteams \
 -e WATCHTOWER_NOTIFICATIONS_MSTEAMS_WEBHOOK_URL \
 =https://webhook.teams.microsoft.com/xxx containrrr/watchtower
```

Utilisation de fichiers de configuration

Comme alternative au passage de variables d'environnement, vous pouvez définir vos configurations de notification dans des fichiers YAML et ainsi les monter dans le conteneur Watchtower. C'est plus « propre » car ainsi vous pourrez conserver vos configurations sans avoir à passer des instructions sans fin. Les chemins de fichiers de configuration à employer sont /config/email.yaml pour la configuration email, /config/slack.yaml pour la configuration Slack et /config/msteams.yaml pour MS Teams. Pour les notifications par e-mail, cela donnera :

```
docker run -d --name watchtower \
 -v /var/run/docker.sock:/var/run/docker.sock \
 -v /path/to/email.yaml:/config/email.yaml \
 containrrr/watchtower
```

Où le fichier email.yaml contiendra :

```
email:
 from: addressexp@gmail.com
 to: addressdest@undomaine.fr
 server: smtp.gmail.com
 port: 587
 user: addressexp@gmail.com
 password: SMTP_password
```

Cela permet de conserver ces configurations de notification à part du conteneur Watchtower et de les réutiliser plus facilement.

BLOQUER LA VÉRIFICATION ET LA MISE À JOUR DE CONTAINERS

Il peut arriver que la nouvelle version d'une image introduise un bug ou une régression, et qu'après la mise à jour le container concerné ne démarre plus. Il est impossible de le détecter puisqu'il s'agit d'erreurs ou de malfaçons non intentionnelles et aléatoires. Néanmoins, vous pouvez exclure les containers sensibles du process, en vue de les mettre à jour individuellement après vérification des retours utilisateurs sur les nouvelles images déployées. Si, donc, vous souhaitez bloquer la vérification et la mise à jour automatique de certains containers sensibles par Watchtower, ajoutez une partie labels dans le fichier docker-compose.yml si elle n'existe pas et à l'intérieur la ligne - "com.centurylinklabs.watchtower.enable=false" :

```
version: "3"
services:
  containerSensible:
    < configuration classique >
    labels:
      - "com.centurylinklabs.watchtower.enable=false"
```

Mise à jour « à la carte » des conteneurs

Watchtower surveille par défaut tous les conteneurs s'exécutant sur votre démon Docker et met à jour tous ceux pour lesquels l'image a été actualisée. Ce comportement est modifiable. Vous pouvez exclure certains conteneurs de la mise à jour via leur nom ou leur étiquette ou n'inclure que certains d'entre eux et exclure tous les autres.

Exclure sur le nom de conteneur

Passez pour cela l'option --exclude avec un filtre regex (expression régulière) pour les noms, comme ceci :

```
docker run -d --name watchtower \
 -v /var/run/docker.sock:/var/run/docker.sock \
 containrrr/watchtower --exclude
"nom_dun_container|nom_dun_autre_container"
```

La barre verticale, vous l'aurez sans doute compris, permet d'associer plusieurs noms. Elle agit donc comme un « et ».

Exclure par étiquette de conteneur

Pour exclure des conteneurs, vous devez les étiqueter avec com.centurylinklabs.watchtower.enable=false, comme ceci pour le conteneur containerlab :

```
docker run -d --label com.centurylinklabs.watchtower.enable=false
containerlab
```

Inclure uniquement des conteneurs spécifiques

Vous pouvez dans l'autre sens créer une « whitelist » de conteneurs à mettre à jour automatiquement en excluant tous les autres via l'option --include :

```
docker run -d --name watchtower \
 -v /var/run/docker.sock:/var/run/docker.sock \
 containrrr/watchtower --include "container-1|container-2|container-3"
```

Modifier les intervalles de vérification

Watchtower vérifiera par défaut l'existence de nouvelles images toutes les cinq minutes. Vous pouvez modifier cette fréquence grâce à l'option --interval et la valeur souhaitée. Attention, si vous ne spécifiez pas d'unité (m pour minutes, h pour hours, etc.), celle par défaut est la seconde :

```
docker run -d --name watchtower \
 -v /var/run/docker.sock:/var/run/docker.sock \
 containrrr/watchtower --interval 120m
```

Cela fera vérifier l'existence de nouvelles images à Watchtower toutes les 2 heures.

Redémarrer les conteneurs

Watchtower attendra par défaut 10 minutes après une mise à jour d'image avant de redémarrer les conteneurs concernés. Vous pouvez modifier cet intervalle de temps avec l'option --restart-delay :

```
docker run -d --name watchtower \
 -v /var/run/docker.sock:/var/run/docker.sock \
 containrrr/watchtower --restart-delay 5m
```

Le délai de redémarrage est fixé ici à 5 minutes.

Configurer Watchtower dans Docker Compose

Watchtower peut aussi être exécuté dans le cadre d'un stack Docker Compose. Ajoutez pour cela un service watchtower à votre fichier docker-compose.yml, comme ceci :

```
version: "3"
services:
 app:
  image: myapp
  ports:
   - "8080:80"
 watchtower:
  image: containrrr/watchtower
  volumes:
   - /var/run/docker.sock:/var/run/docker.sock
  command: --interval 120
```

Cela démarrera Watchtower avec les conteneurs d'application et services définis dans votre fichier compose. Watchtower surveillera les autres services et mettra automatiquement à jour toutes les images lorsque de nouvelles versions seront publiées. Le montage volumes monte le socket Docker afin de permettre à Watchtower de communiquer avec le daemon Docker. L'option command définit l'intervalle de vérification à 120 secondes. Nous avons vu que cet intervalle pouvait être personnalisé selon vos besoins. L'emploi de Docker Compose vous permet d'ajouter facilement la surveillance Watchtower aux piles existantes et futures.

Options de fonctionnement de Watchtower

Pour mettre en place Watchtower, créez un nouveau fichier docker-compose.yml dans un nouveau dossier et ajoutez-y ces quelques lignes :

```
version: "3"
services:
 watchtower:
  image: containrrr/watchtower:latest
  container_name: watchtower
  restart: unless-stopped
  environment:
   - WATCHTOWER_POLL_INTERVAL=300
   - WATCHTOWER_CLEANUP=true
   - WATCHTOWER_INCLUDE_RESTARTING=true
   - WATCHTOWER_LOG_LEVEL=error
   - WATCHTOWER_HTTP_API_METRICS=false
```

Voici à quoi correspondent ces différents éléments de configuration :

- WATCHTOWER_POLL_INTERVAL : c'est l'intervalle de vérification des mises à jour des containers, ici paramétré à 300 secondes (5 minutes).
- WATCHTOWER_CLEANUP : supprime automatiquement les anciennes images inutilisées suite à la mise à jour d'un container lorsqu'il est activé, comme ici, par la valeur true.
- WATCHTOWER_INCLUDE_RESTARTING : redémarre les containers lors de leur mise à jour lorsque sa valeur est à true.
- WATCHTOWER_LOG_LEVEL : le niveau du log d'erreurs. La valeur error le positionne au niveau de détails maximum.
- WATCHTOWER_HTTP_API_METRICS : active ou désactive l'API de metrics qui permet à des logiciels de mesure comme, par exemple, Prometheus de récupérer des informations sur l'exécution de Watchtower.

Avantages et inconvénients

En conclusion, Watchtower est un outil puissant et très pratique pour automatiser la gestion des mises à jour sur Docker. L'ensemble de vos containers peut ainsi rester à jour en permanence, tant pour les fonctionnalités que pour les mises à jour de sécurité, sans que vous ayez à faire quoi que ce soit. Cependant, malgré ces avantages indéniables en termes de maintenance et de gestion des versions, cela peut également entraîner des risques. Le principal inconvénient avec cette méthode est rencontré dans le cas où une nouvelle version d'image introduit un bug, une régression ou même une modification de la configuration qui entrainerait le non-redémarrage du container. Les mises à jour ne sont plus maîtrisées directement par les développeurs et techniciens, mais par un outil d'automatisation tournant à intervalles réguliers. De fait, un service peut donc tomber en panne de manière aléatoire. Il faut donc évaluer avec précision quelle est l'approche la mieux adaptée pour chaque container. L'idéal sera de combiner l'automatisation via Watchtower pour les containers de moindre importance ou ceux pour lesquels les mises à jour sont très fiables avec une supervision manuelle pour les autres, afin de garantir la stabilité et la fiabilité de l'ensemble de votre système. □

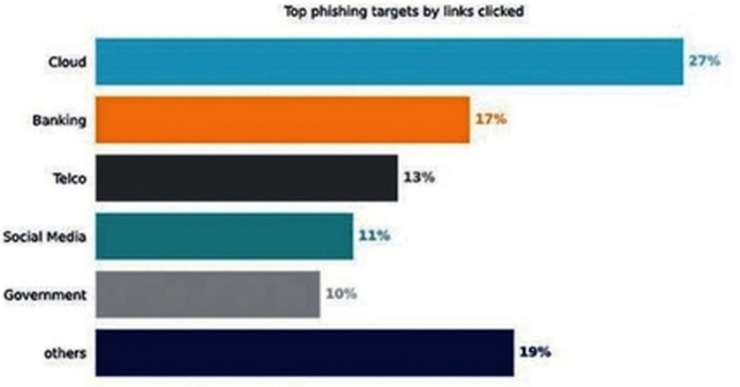
T.T

Cyber

Les utilisateurs moins prudents face au phishing

Un nouveau rapport réalisé pour le compte de Netskope démontre que les utilisateurs ont plus cliqué sur des liens de phishing en 2024 que l'année précédente.

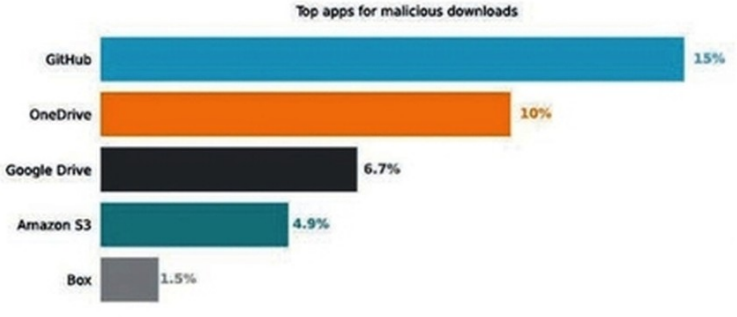
Selon l'étude, les employés des grandes entreprises ont cliqué sur des liens frauduleux près de trois fois plus en 2024 qu'en 2023, en raison de la croissance constante de la quantité et de la qualité des attaques par phishing. Autre enseignement de ce rapport, la hausse des risques de sécurité liés à l'utilisation persistante d'applications cloud personnelles, ainsi qu'à l'adoption continue d'outils d'intelligence artificielle générative (IA générative) dans les environnements professionnels.



Plus de huit utilisateurs sur 1 000 ont cliqué sur un lien frauduleux tous les mois, soit une progression de 190 % par rapport à 2023, année comptant moins de trois utilisateurs sur mille victimes d'une tentative d'hameçonnage. Le lieu d'hébergement des contenus frauduleux fait également partie des éléments d'ingénierie sociale. Les cyberattaquants ciblent des plateformes bénéficiant de la confiance implicite de leurs utilisateurs, notamment de populaires applications cloud telles que GitHub, Microsoft OneDrive ou Google Drive. En 2024, elles furent la source de téléchargements de contenus malveillants au moins une fois par mois dans 88 % des entreprises. Les applications cloud représentent la principale cible des campagnes de phishing sur lesquelles les utilisateurs ont cliqué en 2024, avec plus d'un quart de l'ensemble des clics (27 %). Parmi celles-ci, Microsoft est de loin la marque la plus ciblée (42 %), les attaquants visant les identifiants Microsoft Live et Microsoft 365.

Pro ou perso ?

88 % des employés ont utilisé chaque mois des applications cloud personnelles, tandis que plus d'un quart d'entre eux (26 %) ont téléchargé, publié ou envoyé des données vers des applications personnelles. La fuite de données sensibles par l'intermédiaire d'applications personnelles reste une préoccupation majeure pour la plupart des entreprises, le type de violation des règles de protection le plus courant concernant les données régulées (60 %), ce qui inclut les données personnelles, financières ou de santé téléchargées vers des applis personnelles. Parmi les autres types de données, figurent la propriété intellectuelle (16 %), les codes source (13 %), les mots de passe et les clés (11 %), ainsi que les données chiffrées (1 %).



Et l'IA dans tout ça ?

Alors que les applications d'IA générative ont continué de conforter leur position incontournable (94 % des organisations les utilisent aujourd'hui) en 2024, les entreprises ont montré qu'elles commencent seulement à mettre en œuvre des contrôles afin de garantir leur utilisation sécurisée, et de minimiser les risques liés aux données qu'elles posent. 34 % des organisations se servent d'un coaching utilisateur interactif en temps réel pour renforcer la capacité de leurs collaborateurs à prendre des décisions réfléchies et pertinentes, mais à 73 %, les utilisateurs avertis d'une compromission potentielle de leur entreprise décident de ne pas agir conformément aux informations obtenues lors des séances de coaching. En conséquence, les entreprises ont mis en place des logiciels de DLP (Data Loss ou Leak Prevention) à 45 %. 73 % des entreprises bloquent au moins une application d'IA générative, avec un taux constant de 2,4 outils d'IA générative bloqués en moyenne par an. Le nombre d'applications bloquées a plus que doublé. □

B.G

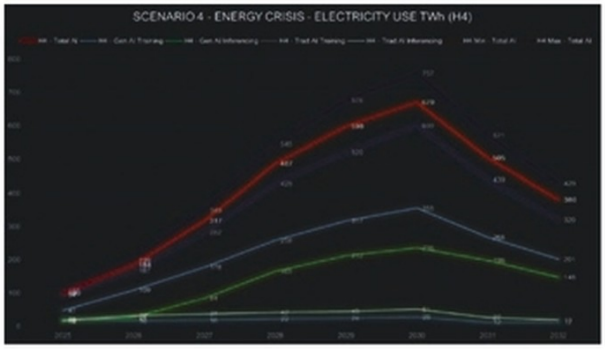
IA De la catastrophe au durable

L'institut de recherche en durabilité de Schneider Electric a rendu publique une étude sur la consommation énergétique de l'IA. Les conclusions de ce rapport évaluent trois scénarios possibles suivant les évolutions à venir en termes d'énergie.

A priori, il ne semble pas qu'un problème urgent puisse toucher les utilisateurs d'intelligence artificielle. Selon le rapport, la demande électrique des centres de données dédiés à l'intelligence artificielle ne représente que 0,3 % de la demande totale d'électricité. C'est sans compter son développement rapide. L'Agence Internationale de l'Energie (AIEA) prévoit que la consommation des centres de données devrait dépasser les 1000 TWh/jour, soit la consommation d'un pays comme le Japon. A la fin de cette décennie, la consommation estimée par le cabinet Gartner serait proche du doublement de la consommation en 2022 pour l'ensemble des centres de données soit, 3 à 4 % de la consommation mondiale d'électricité. Dans certains pays, la consommation atteint déjà des niveaux inquiétants. En Irlande, en 2026, les datacenters consommeront 32 % de l'électricité du pays. En Virginie du Nord, cela représente déjà 25 % de la consommation pour doubler d'ici 2026.

Trois scénarios plausibles

A partir de ce contexte, l'institut a bâti trois scénarios plausibles. Le premier, et le plus optimiste, propose que l'IA elle-même va permettre de réaliser des progrès substantiels sur l'optimisation des ressources et l'efficacité de l'énergie, afin d'améliorer les opérations des centres de données. Cela permettrait de mettre en place un cycle vertueux vers une utilisation durable des ressources énergétiques et une conception plus responsable des équipements.



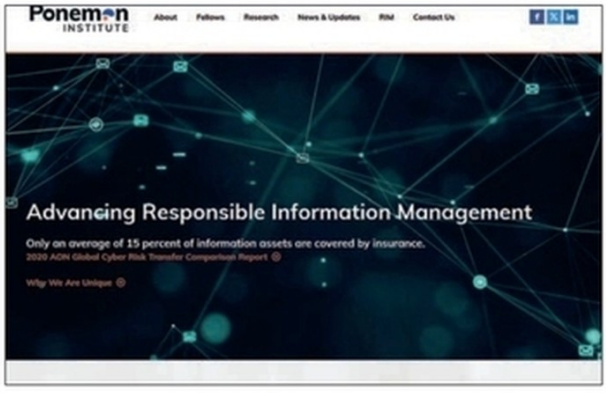
Les conséquences sur l'IA du scénario de crise.

Certains pensent même qu'il faudrait en fait généraliser l'IA partout et dans tous les domaines, afin de permettre des progrès et une utilisation sans limite pour la technologie. Le scénario médian voit des limites et des freins sur le premier scénario provoquant des limitations sur la croissance de l'intelligence artificielle. Enfin, le scénario catastrophe prévoit que le développement de l'intelligence artificielle tel qu'actuellement conduit à une crise énergétique sévère qui entrera en conflit direct avec les autres secteurs critiques de l'économie. Ce scénario s'appuie sur différents éléments corollaires à un développement sans contrôle dans le développement de l'intelligence artificielle : insuffisance de planification du réseau électrique, une prévision de la demande d'IA inexacte... □ **B.G**

Training Comment préparer un audit de conformité informatique

Un audit de conformité informatique est une analyse indépendante des outils pratiques et politiques de cybersécurité d'un organisme. Son but est de s'assurer que votre organisation respecte les réglementations et lois spécifiques déterminées par les organismes de certification et autres autorités habilitées. Nous allons voir dans cet article comment préparer au mieux ce processus essentiel.

La réussite d'un audit signifie plusieurs choses. D'abord que vous avez mis en œuvre les meilleures stratégies de cybersécurité, afin de protéger vos données sensibles et d'atténuer au maximum les risques liés à la sécurité. Ensuite que vous avez donné la priorité à la confidentialité de toutes les parties prenantes, ceci incluant vos clients et vos éventuels investisseurs. Enfin ,que vous avez économisé de potentielles amendes pour non-conformité. D'après une étude de l'Institut Ponemon, les pénalités pour non-respect des règlements sur la protection des données (RGPD, IA Act &Co) coûterait, en moyenne, deux fois plus cher que le maintien de la conformité. Le calcul est vite fait, même pour les entreprises les moins consciencieuses.



L'Institut Ponemon est spécialisé dans la recherche et la formation sur le thème de l'utilisation responsable de l'information et des problèmes critiques en matière de sécurisation des données et des infrastructures IT.

Voici quelques recommandations simples qui vous aideront à atteindre cet objectif :

- Informez les employés sur tous les aspects inhérents à la confidentialité des données, et donnez-leur les bons outils pour en protéger l'accès.
- Fournissez aux employés mobiles des ordinateurs portables et des appareils intégrant des politiques de sécurité et des mécanismes de prévention sérieux et efficaces, ainsi qu'un accès sécurisé aux données de l'entreprise.
- Mettez en place des mécanismes de contrôle d'autorisation, afin de limiter l'accès aux installations d'applications. N'autorisez que les logiciels et applications approuvés. L'idéal est de créer un store d'entreprise et de bloquer tout le reste pour le commun des utilisateurs.
- Utilisez prioritairement des solutions de stockage dans le cloud sécurisées et modernes.

Passer un audit de conformité informatique ne doit pas devenir un calvaire. Avec la bonne stratégie et une bonne organisation, vous pourrez préparer en toute confiance votre audit et exécuter ce processus sans heurt du début à la fin. Voyons quelles en sont les étapes clés.

CONTRÔLER EN PERMANENCE LA CONFORMITÉ

La conformité est un processus continu et non figé. Les organismes de réglementation peuvent mettre à jour leurs règles et de nouvelles technologies de cybersécurité évoluent à un rythme rapide pour s'adapter aux menaces. Pour rester à jour, vous devez donc surveiller en permanence vos contrôles informatiques et vos processus système. Suivez ces recommandations afin de maintenir la conformité réglementaire :

- **Mettez en place des outils de gestion de la conformité pour suivre et rapporter en permanence le statut de conformité et les problèmes potentiels**
- **Abonnez-vous aux bulletins d'information du secteur, suivez les mises à jour réglementaires et participez aux webinaires pertinents**
- **Consultez des experts en conformité et en cybersécurité**
- **Maintenez tous les logiciels et systèmes à jour avec les derniers correctifs**

Identifier les exigences réglementaires spécifiques

La première étape consiste à comprendre quelle réglementation s'applique à votre organisation. Chaque secteur d'activité est régi par des réglementations et des organismes différents. La loi HIPAA (Health Insurance Portability and Accountability Act), par exemple, est celle applicable dans le secteur de la santé, tandis que la norme PCI-DSS (Payment Card Industry Data Security Standard) concerne le commerce de détail. Le secteur des services financiers est, lui, régi par la loi SOX (Sarbanes-Oxley) pour les

rapports financiers et la loi GLBA (Gramm-Leach-Bliley) pour la protection des informations sur les clients. Cette phase est essentielle car si les normes adéquates ne sont pas identifiées, vos efforts de mise en conformité seront incomplets et donc inutiles. Il faut par conséquent absolument rechercher les lois et règlements spécifiques à votre secteur d'activité. La première chose à faire pour cela est une recherche en ligne, en vue de récupérer des informations souvent précieuses. Cependant, cela ne suffira pas. Il est indispensable de consulter des experts tels que des avocats ou des consultants en conformité. Il existe des cabinets de conformité qui pourront vous aider à vous y retrouver dans cette jungle de réglementations.

Établir une communication claire avec l'équipe d'audit

Pendant l'audit, le maintien d'une communication claire entre l'équipe interne et les auditeurs est essentiel. Cela permet à chacun de comprendre les objectifs et les attentes de l'audit afin d'éviter les retards et les malentendus. Une communication efficace permet à l'équipe d'audit d'aborder rapidement toute préoccupation ou tout problème de non-conformité, afin de les résoudre avant qu'ils ne deviennent plus importants. Elle instaure un climat de confiance entre l'entreprise et les auditeurs, favorisant ainsi un environnement transparent. Cet aspect est particulièrement important lorsqu'il s'agit de traiter avec des organismes de réglementation, car il témoigne de votre engagement en faveur de la conformité et contribue à asseoir votre crédibilité. Au cours de ces étapes, les auditeurs de conformité s'entretiendront avec les parties prenantes de chaque service, afin d'obtenir une description précise de vos processus informatiques. Vous devrez veiller à bien préparer vos équipes informatiques à collaborer avec les auditeurs en répondant à leurs questions et en leur fournissant un accès facilité à vos systèmes. Si, par exemple, vous travaillez dans le commerce de détail, vérifiez que les auditeurs aient bien accès à tous vos grands livres et journaux de paiement. Cela les aidera à vérifier votre conformité à la norme PCI-DSS et à s'assurer que rien n'a été oublié.

Vous trouverez, sur Internet, nombre de sites qui vous conseilleront sur la réalisation de tel ou tel audit — en vous proposant bien entendu leurs services. Le site de secureframe, par exemple, vous donne des informations sur la réalisation d'un audit interne ISO 27001.

Les différents types d'audits de conformité informatique

Les audits de conformité informatique varient en fonction de leur objectif et de leurs buts respectifs. Nous allons voir quels sont les principaux types existants pour maintenir votre organisation sur le bon chemin, loin des amendes et des risques divers.

Audit interne versus audit externe

Les audits internes sont menés par votre propre équipe ou par des auditeurs internes, et se concentrent sur l'évaluation de l'efficacité des contrôles, processus et systèmes internes de votre organisation. Ces audits sont continus et vous aident à identifier et à traiter les problèmes avant qu'ils ne s'aggravent. Les audits externes, quant à eux, sont réalisés par des tiers indépendants. Ils fournissent une évaluation objective de votre conformité aux exigences réglementaires et aux normes industrielles. Ces audits sont souvent exigés par les organismes de réglementation ou les parties prenantes, et offrent une perspective nouvelle sur le statut de conformité de votre organisation.

Audits financiers dans l'informatique

Les audits financiers sont un élément clé d'une gestion efficace de l'information pour une bonne gouvernance des données. Ces audits ciblent spécifiquement l'exactitude et l'intégrité des informations et des processus financiers gérés par des systèmes informatiques. Ils examinent la manière dont les transactions financières sont enregistrées, traitées et restituées. Ils garantissent que les données financières sont fiables et que des contrôles internes sont en place pour prévenir les fraudes et les erreurs.

Audits des contrôles des systèmes et de l'organisation (SOC)

Les audits SOC se concentrent sur l'évaluation des contrôles et des processus qui ont un impact sur la fiabilité des rapports financiers et la sécurité des données. Les audits SOC 1 évaluent les contrôles liés aux rapports financiers. Les audits SOC 2 évaluent les contrôles liés à la sécurité, la disponibilité, l'intégrité du traitement, la confidentialité et la protection de la vie privée. SOC 3 fournit un aperçu général de la sécurité des données SOC 2.

Audits conformes à la norme ISO/IEC 27001

Il s'agit d'audits externes réalisés sur la base des publications de conformité de l'ISO/IEC 27001. L'ISO (Organisation internationale de normalisation) et la CEI (Commission électrotechnique internationale) évaluent le degré de préparation de votre entreprise contre les risques liés aux violations de données, au piratage et aux fuites d'informations. Cette norme, reconnue au niveau international, porte sur l'établissement, la mise en œuvre, la maintenance et l'amélioration continue d'un système de management de la sécurité de l'information (SMSI). La GRC (gouvernance, risque et conformité) est une stratégie intégrée permettant de gérer efficacement et correctement les politiques, les processus et les contrôles. ☐ **T.T**

Salaires

DSI, IA, data, cloud et cyber, le quinté gagnant des profils les plus recherchés

La pénurie dans les métiers digitaux et IT est globalement moins forte que les années précédentes, entraînant une hausse de salaire moyenne de « seulement » 4 % en 2024. Elle reste conséquente dans certains métiers liés à l'IA, aux données, au cloud et à la cybersécurité, ainsi que dans les postes de direction.

La tendance à la stabilisation générale des salaires sur le marché de l'emploi tech a été observée en 2024, avec une hausse moyenne de salaire médian brut de 4 % contre 7 % en 2023, selon le benchmark annuel sur les salaires du cabinet de recrutement Aravati publié fin 2024. Baptisé « HR Pulse », il repose sur la base des déclarations de 27 900 candidats avec lesquels le cabinet a échangé pendant un an. Cette tendance devrait se poursuivre en 2025, selon Hymane Ben Aoun, directrice générale d'Humanskills, qui rassemble des entreprises de services RH (dont Aravati qu'elle a cofondé) : *« La fête est finie côté salaires, même s'il y a des augmentations encore élevées ciblées sur certains postes. Les salaires devraient se stabiliser en 2025. Ce sont les entreprises qui détiennent le pouvoir de négociation. Elles se montrent plus prudentes et moins pressées qu'auparavant, et mettent plus de temps à recruter. Quatre facteurs l'expliquent. Il y a aujourd'hui beaucoup de candidats sur le marché, donc plus forcément besoin de débaucher des personnes en poste, par exemple des développeurs full stack. Les entreprises se préparent progressivement à appliquer la directive européenne sur la transparence des salaires, qui doit être transposée en droit français d'ici juin 2026. Les exceptions à la grille des salaires sur les métiers en tension devraient donc diminuer. Et le trou d'air dans les levées de fonds des startups, notable depuis 2023, les font se séparer de collaborateurs jugés trop chers, notamment dans le développement, remettant des candidats sur le marché. Enfin, les incertitudes économiques mondiales, dues au contexte géopolitique et à l'inflation, jouent sur la croissance et les investissements des grandes entreprises. »*

L'étude montre quels sont les profils les plus recherchés en 2025, donc les plus difficiles à recruter, avec des augmentations et niveaux de salaire qui resteront élevés. Le marché de l'emploi tech reste une fête pour les top managers, les experts en IA, data, cloud et cyber.

Dirigeants techniques : +13 % de salaire en moyenne

Avec une hausse moyenne de rémunération de 13 % en 2024, les postes de direction à forte composante technique sont jugés stratégiques, dans un environnement technologique en constante évolution : DSI, CISO (chief information security officer), head of AI. Avec 10 à 15 ans d'expérience, les CIO et CTO ont une rémunération moyenne de 140 K€, celles et ceux ayant plus de bouteille peuvent atteindre 250 K€. *« Le DSI reprend une dimension hautement stratégique, avec un périmètre de responsabilités étendu,* note Hymane Ben Aoun. *Il est responsable de la vision digitale globale, sachant doser les niveaux d'innovation et de sécurité nécessaires à chaque projet. Avec l'essor des technologies comme l'IA, la cybersécurité, l'automatisation et la gestion des données massives, il joue un rôle clé dans le pilotage de l'intégration des nouvelles technologies, et dans l'alignement des systèmes sur les objectifs stratégiques. Il devient aussi le principal architecte des plateformes numériques qui façonnent l'expérience client et l'efficacité opérationnelle. Notons l'apparition du CAIO, rattaché au comité exécutif ou au DSI, qui pilote l'intégration de l'IA, dans les entreprises où l'IA est une clé de la transformation digitale. »*

Salaire médian brut annuel dans la data — NIVEAU D'EXPÉRIENCE →

	0 - 3 ans	3 - 5 ans	5 - 10 ans	+10 ans
Responsable de la Connaissance Clients	-	-	65-80 K€	80-100 K€
Responsable de la Performance Digitale	-	55-65 K€	65-80 K€	80-120 K€
Responsable BI	-	60-75 K€	75-100 K€	100-145 K€
Data Quality Manager	38-45 K€	45 -55 K€	55-70 K€	-
Responsable Data	-	65 -80 K€	80-95 K€	95-130 K€
Architecte Data	-	60-75 K€	75-100 K€	-
Data strategy consultant	-	-	75-100 K€	100-130 K€
Data Scientist	45 -65 K€	60-80 K€	80-95 K€	100-150 K€
Dataminer	-	-	60-75 K€	-
Data Analyst / Business Analyst	40-50 K€	50 - 60 K€	60-90 K€	-
Data Steward	40-45 K€	45 -60 K€	60-70 K€	-
Data Engineer	40-50 K€	50 -75 K€	80-100 K€	-
DataOps	40-50 K€	50-75 K€	+75 K€	-

Source : HR Pulse — Benchmark des salaires 2025 — Aravati

Salaire médian brut annuel dans l'IA (niveau d'expérience)	0 - 3 ans	3 - 5 ans	5 - 10 ans	+10 ans
Machine learning engineer	50-70 K€	70-90 K€	90-120 K€	120-160 K€
Deep learning engineer	55-75 k€	75-95 K€	95-130 K€	130-190 K€
AI Research scientist	55-75 K€	75-100 K€	100-130 K€	130-180 K€
AI solutions architect	-	90-110 K€	110-140 K€	140-190 K€
AI consultant	50-70 K€	70-90 K€	90-120 K€	120-160 K€
NLP engineer	55-75 K€	75-95 K€	95-125 K€	125-170 K€
Robotics engineer	50-70 K€	70-90 K€	90-120 K€	120-160 K€
Computer vision engineer	50-70 K€	70-90 K€	90-120 K€	120-160 K€
AI operations engineer	50-70 K€	70-90 K€	100-130 K€	130-170 K€
Machine Learning Ops	45-60 K€	60-80 K€	80-110 K€	110-150 K€
AI security Specialist	60-80 K€	80-100 K€	100-130 K€	130-170 K€
Spécialiste en éthique de l'IA	50-70 K€	70-90 K€	90-120 K€	-
Coach en IA	-	55-70 K€	70-90 K€	90-120 K€
Chief IA officer	-	-	130-180 K€	180-230 K€

Source : HR Pulse — Benchmark des salaires 2025 — Aravati

Dans ce contexte pénurique, la demande en managers de transition a cru en 2024 de 20 %, du fait de la nécessaire poursuite des programmes de transformation avec des budgets RH plus contraints, et de l'accélération des projets dans les ETI et en B2B. Si la demande de managers de transition est essentiellement francilienne, un tiers des missions a été réalisé en région en 2024.

Gros besoins en experts IA et data

La pénurie de profils spécialisés en IA s'explique par une augmentation des besoins de profils spécialisés pour développer son usage au sein des différents départements (marketing, supply chain, RH, opérations...). Les entreprises sont à l'affût d'ingénieurs LLM Engineers, data scientists, AI ethics & compliance officers... Ainsi, les experts en IA, demeurent très recherchés, ont vu leur salaire augmenter de 12 % en 2024 (dépassant les 90 K€ pour 5 à 10 ans d'expérience), une hausse similaire à celle de 2023. Les Machine Learning Ops, chargés de déployer les algorithmes en production, émergent avec des salaires atteignant 80 K€ pour plus de 5 ans d'expérience.

Avec l'accélération des besoins de gestion de produits data complexe, intégrant l'IA, les product owners data et data product managers seront fortement sollicités en 2025. Data engineers et data analysts restent également très demandés, avec une hausse de salaire moyenne de 6 % en 2024.

Les profils IT spécialisés en tension

En 2024, le recrutement de profils techniques a été marqué par une pénurie de talents spécialisés, une forte demande pour les compétences en intelligence artificielle, cloud computing et cybersécurité ; et une compétition accrue entre entreprises pour attirer les meilleurs experts techniques. La rémunération des développeurs est variable : plus la stack est demandée et les profils peu nombreux, plus les enchères montent. Côté infrastructure, systèmes et réseaux, le marché est moins tendu pour les postes les plus classiques, tandis que les ingénieurs cloud et les experts en cybersécurité — tels les pentesters — seront très sollicités en 2025. DevOps et ingénieurs de fiabilité de site (SRE) vont être particulièrement recherchés.

Evolution des métiers design et produit

« En design, les entreprises souhaitent de la polyvalence, remarque Hymane Ben Aoun, *préférant embaucher des system designers pour remplacer des UX/UI designers, pour créer des systèmes cohérents et durables. »* Le UX strategist gagne en importance, intégrant une vision globale de l'expérience utilisateur.

Demande et salaires restent en hausse dans les métiers du product management, du fait de leur rôle de plus en plus stratégique et de la demande de compétences complémentaires fonctionnelles et techniques. Le salaire du Product Ops, essentiel dans la coordination et l'optimisation des processus produit, peut atteindre 80 à 100 K€ pour les profils les plus expérimentés.

Le marché de la communication et du marketing digital est quant à lui morose, avec des rémunérations stables (+2 % par rapport à 2023).

Demande de compétences pointues

« Les entreprises demandent en général des profils plus experts, plus impliqués dans l'amélioration des systèmes, conclut Hymane Ben Aoun. *Nous n'avons jamais eu autant de candidatures spontanées. Mais le marché regorge de candidats avec des formations un peu légères, notre rôle est de creuser afin de trouver celles et ceux qui ont les compétences demandées. Nous remarquons également que la demande des entreprises depuis mi-décembre repart globalement à la hausse après un 2e semestre 2024 plutôt calme. »*

Les compétences métier pointues sur les technologies les plus récentes sont les plus demandées. Les métiers dont les tâches vont de plus en plus être automatisées et assistées par l'IA, comme ceux du design, de la communication et du marketing digital, ou même dans le développement, marqué par l'essor du no-code et du low-code, devront s'adapter pour que les entreprises continuent à leur faire les yeux doux. La montée en compétences, voire la reconversion, grâce à la formation continue, est un enjeu clé tant pour les candidats que pour les entreprises à la recherche de profils en pénurie. La montée en compétences concerne aussi bien les compétences techniques que les soft-skills, de mieux en mieux valorisées car elles permettent une adaptation plus aisée aux changements sociétaux, technologiques et organisationnels auxquels font face les entreprises.

C.C

Cybersécurité 2025 va secouer

Sommaire

Au risque de se répéter, peu de chances que le risque cyber désenfle en 2025. Deepfakes, attaques par force brute, quishing... d'une année sur l'autre, la menace est toujours plus protéiforme, tandis que des enjeux complexes, comme la cryptographie post-quantique, pointent à l'horizon.

Et que dire de l'IA ? Elle alimente toujours autant les craintes que les espoirs. Génératrice de nouvelles attaques ou rempart défensif, elle pousse la cybersécurité dans une bataille permanente entre innovation et vulnérabilité. Les observateurs oscillent donc entre crainte et optimisme. Crainte de voir les attaques dopées à l'IA, souvent spectaculaires et très médiatisées, se généraliser ; optimisme, car 2024 nous a appris que cette technologie était surtout profitable aux défenseurs pour le moment.

Et que dire du changement majeur qui s'est opéré de l'autre côté de l'Atlantique ? L'arrivée de Donald Trump au pouvoir marquera, et marque déjà, une politique économique tournée vers la compétitivité des États-Unis avant tout. Ce qui ne sera pas sans incidence sur la cybersécurité ou encore la protection des données.

Reste qu'en Europe, pour beaucoup d'acteurs économiques, 2025 sera placée sous le signe de la mise en conformité avec DORA et NIS2, afin d'améliorer leurs pratiques de cybersécurité et leur résilience (la fameuse), avec toute l'agilité comptable et l'expertise juridique que cela nécessite.

Les évolutions cyber attendues pour 2025

Au cœur des enjeux cyber en 2025, les nouvelles menaces générées par l'intelligence artificielle (IA) ne doivent pas occulter les autres risques, qui exigent des éditeurs d'innover, des entreprises de s'armer toujours plus et toujours mieux, et des collaborateurs de prendre la mesure du risque.

Comme chaque année, les pythies de la sphère cyber nous livrent leurs oracles pour 2025. Difficile d'y couper : l'intelligence artificielle est au cœur des enjeux. Comme nous l'expliquions dans notre dossier sur le bilan 2024 (voir numéro 232 de l'INFOCR), les attaques alimentées par l'IA ne se sont pas généralisées comme le craignaient certains observateurs, et cette technologie reste avant tout une affaire de défenseurs. Une tendance qui se confirmera en 2025, tant que les cybercriminels n'auront pas les ressources nécessaires pour l'exploiter pleinement. Contrairement aux éditeurs, MSP et aux autres entreprises de cybersécurité qui l'intègrent à leurs outils pour améliorer leurs capacités de détection et de remédiation.

Bien qu'encore sporadique, la menace n'en est pas moins réelle. Et si les méthodes plus traditionnelles et éprouvées continuent d'être largement employées, oui, les escroqueries de type deepfakes, la génération de code malveillant, les e-mails de phishing dopés à GenAI, ou encore les attaques par force brute exploitant les capacités de l'IA vont s'accélérer en 2025.

L'ère des sous-doués de la cybercriminalité

Dans son bulletin de sécurité annuel 2024, la Global Research and Analysis Team (GReAT) est revenue sur son travail de veille concernant 900 groupes APT (menaces persistantes avancées) et a constaté un recours récurrent aux solutions d'IA. Les chercheurs citent, par exemple, le groupe Lazarus, qui a utilisé des images générées pour exploiter une vulnérabilité zero-day de Chrome et voler des cryptomonnaies. D'autres groupes ont distribué des versions malveillantes et détournées de modèles d'IA open source dans lesquelles était injecté du code malveillant.

Le GReAT craint qu'à terme, les LLM (modèles de langage de grande taille) ne deviennent « des outils standard » qui épauleront les cybercriminels dans leur travail de reconnaissance, de détection des vulnérabilités et de génération de scripts malveillants, le tout avec une facilité déconcertante.

Car si développer, exploiter et maintenir des technologies d'IA mobilise des compétences techniques avancées, Sysdig estime toutefois qu'elles rendent la cybercriminalité plus accessible que jamais. Grâce aux LLM, les attaquants n'auront plus besoin de maîtriser des compétences avancées pour mettre sur pied leurs campagnes d'ingénierie sociale, développer des logiciels malveillants ou lancer des attaques complexes.

Un récent exemple tend à donner raison à Kaspersky et Sysdig. L'équipe de recherche de CheckPoint a détecté, en décembre 2024, un ransomware-as-a-service baptisé FunkSec, dont les membres, jugés inexpérimentés, ont développé leurs outils à l'aide de l'IA. *« De manière surprenante, nos découvertes indiquent que le développement des outils du groupe, y compris l'encryptage, a probablement été assisté par une intelligence artificielle, ce qui pourrait avoir contribué à leur itération rapide, malgré le manque apparent d'expertise technique de leur auteur »*, écrivent les chercheurs.

Shadow IT : mieux superviser les IA publiques

Pour 2025, Mimecast attire l'attention sur un autre risque émergent qui a beaucoup fait parler en 2024 : le Shadow IT. Ce phénomène est lié à l'utilisation, par des employés, d'outils d'IA générative publics dans un cadre non approuvé par la société. Cette pratique peut exposer des informations confidentielles qui auraient été utilisées pour produire des résultats. Mimecast met ainsi en garde contre cette pratique, qui peut potentiellement mener à des violations ou à l'exposition de données sensibles restituées dans le cadre d'autres conversations avec un chatbot. Pour 2025, les organisations devront donc continuer à restreindre l'utilisation de ces solutions en les bloquant par défaut, du moins sur les réseaux supervisés. Pour ceux qui ne le sont pas, il faudra avant tout sensibiliser les collaborateurs aux bonnes pratiques et usages, et s'assurer qu'ils utilisent les modèles autorisés en interne.

« Mais on sait qu'intégrer des agents d'IA coûte cher. Et il ne sera pas possible d'endiguer le fait que leurs pendants publics vont être de plus en plus utilisés en Shadow IT », reconnaît Sébastien Baron, directeur technique de Mimecast France. Il préconise plutôt de sensibiliser les entreprises et de les inviter à bloquer certains usages, mais aussi de mettre à la disposition des équipes de sécurité, des moyens de contrôler ce qui est transmis vers les IA publiques pour bloquer et alerter en cas de pratiques contraires aux politiques internes de l'entreprise. *« Il est possible de positionner un agent sur le poste de travail du salarié, qui va intercepter les accès aux fichiers, connaître leurs métadonnées… et déterminer si ce document peut voyager vers les IA publiques type ChatGPT, ou non »*, développe Sébastien Baron.

Apprendre à repérer les deepfakes

Autre risque très médiatisé qui devrait prendre de l'ampleur en 2025 : les deepfakes. Il sera essentiel pour les entreprises de s'armer de vigilance autant que de technologies avancées, et de revoir leurs pratiques pour contrer cette nouvelle forme d'ingénierie sociale. Concrètement, les organisations, les RSSI et les collaborateurs peuvent déjà s'approprier des outils — tels que FakeCatcher, Deepware ou Sensity AI —, afin de mieux détecter les contenus générés par IA en s'appuyant sur des analyses de distorsions, de mouvements incohérents ou de qualité audio suspecte.

À noter que des standards techniques, comme le projet C2PA, sont en cours d'élaboration et auront pour mission de faciliter la traçabilité des contenus afin de garantir leur authenticité.
Les entreprises devront également former leurs employés à repérer les signes et à vérifier la légitimité des contenus, via des processus de validation stricts : rappeler un contact connu pour confirmer une information, appliquer une approche Zero Trust dans le cas d'un appel téléphonique ou d'une visioconférence demandant d'effectuer un virement ou une transaction. Cela inclut l'intégration de vérifications supplémentaires, telles que l'authentification multifactorielle ou la biométrie comportementale.

2025 : une année de la transition vers la cryptographie post-quantique ?

Pour Keyfactor, aucun doute que la transition vers la cryptographie postquantique marquera un tournant en 2025. L'année dernière, le National Institute of Standards and Technology (Nist) a publié les premiers standards d'algorithmes de chiffrement postquantiques résistant aux ordinateurs quantiques capables de casser les algorithmes cryptographiques actuels et de mettre en péril la confidentialité des données. Le Nist enjoint les experts en cybersécurité des entreprises à anticiper et à les intégrer dans leurs systèmes actuels et futurs.
Mais cette transition complexe demande de mener un important inventaire et de prioriser les systèmes à migrer. Si la disponibilité d'ordinateurs quantiques n'est pas envisagée avant 2030, selon les estimations, la plateforme de gestion des identités Keyfactor et d'autres acteurs comme SandboxAQ, spécialisée dans divers domaines du quantique et qui a participé à la soumission de nombreux candidats au Nist, invitent à ne pas perdre de temps pour adopter les nouveaux standards.
« Les entreprises devront évaluer la maturité de leurs postures PQC et l'hygiène de sécurité de leur infrastructure à clé publique (PKI). Attendre la transition n'est plus une option, la date butoir de migration, fixée à 2035 par le Nist — date à laquelle RSA, ECDSA, EdDSA, DH et ECDH seront totalement interdits — laisse peu de place à la procrastination », prévient Keyfactor.

QR codes malveillants : plus c'est simple, plus ça marche

Il y aurait encore beaucoup à dire sur l'IA, mais il ne faudrait pas occulter le reste. Si les attaques 100 % drivées par l'IA sont encore de l'ordre de la science-fiction, Mandiant rappelle que les principales perturbations en 2025 viendront d'attaques plus traditionnelles, comme les ransomwares, qu'elles soient renforcées par l'IA ou pas. La filiale de Google prévoit également que les InfoStealers deviendront un vecteur principal des violations de données. Ces malwares peu sophistiqués permettent aux cybercriminels, même les moins expérimentés, de voler des identifiants. Une efficacité décuplée par le fait que, selon Mandiant, encore trop de collaborateurs en entreprise n'appliquent pas les méthodes basiques d'authentification, comme la 2FA/MFA.
Selon leurs conclusions, le temps d'exploitation des vulnérabilités après leur divulgation devrait continuer de diminuer, avec un temps moyen d'exploitation passant à 5 jours, contre 32 auparavant. *« Cette tendance vers des exploitations plus rapides, notamment des vulnérabilités n-day et zero-day, se poursuivra. En parallèle, le nombre de fournisseurs ciblés continuera d'augmenter, nécessitant une vigilance accrue quant aux surfaces d'attaque. »*
Mimecast met en garde contre le quishing : une méthode de phishing utilisant des QR codes malveillants. Ce risque est encore largement ignoré dans les stratégies des équipes de sécurité — et c'est bien dommage. Entre le 1er avril et le 5 juin 2024, Mimecast dit avoir détecté 70 000 messages malveillants intégrant des QR codes dans des pièces jointes. *« Une tendance qui a doublé par rapport à l'année dernière »*, indique Sébastien Baron. Ces messages visaient à tromper les utilisateurs pour qu'ils fournissent leurs identifiants ou installent des logiciels à leur insu.
Le quishing est particulièrement sournois, car le QR code ne peut être intercepté que par une solution de protection de messagerie. *« Si ce QR code est envoyé dans le corps d'un e-mail ou en pièce jointe, une fois reçu, il sera généralement scanné avec un téléphone, et non un ordinateur d'entreprise qui est équipé de solutions de sécurité supplémentaires »* capables de détecter les compromissions. Avec le QR code, si l'attaque parvient jusqu'à l'utilisateur, il n'existe plus de barrières pour la stopper, contrairement aux URL. *« Les cybercriminels ont bien compris cet avantage »*, prévient Sébastien Baron.
CheckPoint a également constaté une augmentation de ce vecteur dans les attaques de phishing. Son équipe de recherche a repéré une escroquerie récente, consistant à envoyer des e-mails contenant de faux QR codes redirigeant vers des pages frauduleuses de dons pour les victimes des incendies de Los Angeles (États-Unis).
« Lors de catastrophes naturelles ou à l'occasion d'événements d'ordre géopolitique ou culturel, on observe l'émergence de cybercriminels sans scrupules. Ils exploitent la situation en développant de fausses applications dans le but de collecter des données personnelles », confirme David Grout, chief technical officer (CTO) chez Mandiant pour la zone EMEA. Et cette tendance ne devrait pas faiblir en 2025 et pourrait même devenir « systématique ».
Comment se prémunir ? Il existe des solutions spécifiquement conçues pour filtrer et scanner les e-mails en temps réel, afin d'analyser les QR codes avant qu'ils ne soient ouverts.

En parallèle, Mimecast invite les entreprises à mettre en place des formations de sécurité axées sur ce risque, afin de développer des automatismes chez les salariés, comme la vérification des sources avant ouverture et des URL après numérisation.

Une formation qui doit (vraiment) former

La sensibilisation des collaborateurs est souvent présentée comme essentielle pour lutter contre la cybermenace et faire émerger une culture de la cybersécurité en entreprise. Et pour cause, d'après une étude d'IBM, l'erreur humaine est en cause dans 95 % des cyberattaques. Se pose alors la question : les formations proposées sont-elles seulement efficaces ? À l'occasion d'une étude interne menée auprès de 42 000 clients, Mimecast a comparé le nombre d'incidents supplémentaires interceptés entre ceux dotés d'initiatives de sensibilisation (y compris celles de Mimecast) et ceux qui n'en utilisaient pas. Résultat : le taux de réduction d'incidents supplémentaires chez ceux ayant déployé des initiatives d'Awareness Training n'était que de 0,00015 %. Autrement dit, parfaitement inefficaces.

Mimecast prédit ainsi qu'en 2025, les formations devraient évoluer vers des formats moins généralistes, mais en temps réel et intégrées dans les outils de collaboration, afin de mieux cibler les erreurs individuelles de chaque utilisateur. Cela passerait par l'analyse des remontées de différentes solutions de cybersécurité, pour une approche au cas par cas, réorientant les salariés vers des modules de sensibilisation au plus proche de leurs besoins.

Dans les faits, le mouvement semble déjà enclenché, et pas uniquement chez Mimecast. CheckPoint propose une approche multiniveau qui combine formation, simulation de phishing, reporting et contenus adaptés aux comportements des employés. HornetSecurity, de son côté, a développé une offre reposant sur une nouvelle approche plus ciblée : le Security Awareness Service, qui s'adapte à chaque utilisateur grâce à un profilage personnalisé.

Obligation réglementaire et souplesse budgétaire

À l'échelle française et européenne, cette année sera également placée sous le signe d'un durcissement réglementaire, notamment avec l'entrée en application de NIS2 (Network and Information Security) et de Dora (Digital Operational Resilience Act), deux cadres juridiques de l'Union européenne visant à renforcer la cybersécurité et la résilience des systèmes numériques dans des secteurs critiques (voir dossier de L'Informaticien dans ce même numéro). Les chiffres tendent à montrer qu'il y a urgence, et c'est un euphémisme ! Car — il faut le rappeler —, la question pour une organisation n'est pas de savoir si elle va être attaquée, mais quand. D'après un rapport de Cohesity datant de 2024, *« 86 % des décideurs français interrogés ont confirmé que leur entreprise avait été victime d'une attaque par ransomware en 2024, alors qu'ils n'étaient que 53 % l'année précédente »*, et *« 92 % admettent avoir payé une rançon au cours de l'année écoulée »*.

Preuve qu'en matière de cyber-résilience, la route est encore longue. Souvent, pour pouvoir redémarrer leurs opérations et limiter la casse, les victimes estiment ne pas avoir d'autre choix que de passer à la caisse.

Nouvelle percée pour les passkeys

Andrew Shikiar, directeur général de Fido Alliance

Les passkeys sont fortement poussées par la Fido Alliance (Fast Identity Online), une association industrielle qui regroupe Google, Cisco ou encore Apple, et travaille au développement de dispositifs d'authentification sécurisés et interopérables orientés vers le passwordless. Et la Fido Alliance estime que d'ici à la fin 2025, les 1 000 principaux sites web mondiaux proposeront les passkeys. La marche est haute. *« En mai 2024, ils étaient pris en charge par 20 % des 100 principaux sites web mondiaux et 12 % des 250 premiers. »* Mais la technologie, elle, progresse peu à peu. Le Nist (National Institute of Standards and Technology) a actualisé ses directives relatives à l'identité numérique, en reconnaissant que les identifiants synchronisés répondent aux exigences du niveau AAL2. *« La technologie bénéficie déjà d'un élan grâce au soutien des grandes entreprises comme Apple, Google et Microsoft, en raison de ses avantages en matière d'ergonomie et de sécurité, ce qui en fait une alternative attrayante aux mots de passe traditionnels »*, assure Andrew Shikiar, directeur général de Fido Alliance. En Europe, Fido parie sur des directives et incitations de l'Union européenne pour accélérer encore cette tendance. Des initiatives comme l'identité numérique européenne (Eudi), accessible à 448 millions de citoyens d'ici à 2026, ainsi que des réglementations telles que PSD3, devraient promouvoir leur utilisation, notamment dans les paiements où elles pourraient remplacer l'authentification en deux étapes.

En 2025, la cyber-résilience passera aussi par une intégration des contrôles de sécurité aux services informatiques, afin de limiter les frictions entre l'InfoSec et l'IT, et faire en sorte que la sécurité informatique figure parmi les priorités, estime Cohesity. *« Cette transition entraînera des changements importants dans le rôle du RSSI (responsable sécurité des systèmes d'information), la responsabilité du travail quotidien requis pour construire et maintenir la cyber-résilience étant transférée au DSI (directeur des systèmes d'information). Cela conduira à l'absorption d'un plus grand nombre d'équipes RSSI au sein des fonctions informatiques. »*

Quid du poids de ces réglementations sur les budgets en 2025 ?

Selon une enquête menée par Veeam Software et présentée en novembre 2024, 95 % des entreprises ont déclaré avoir réaffecté certains budgets pour financer leur mise en conformité avec NIS2. L'étude suppose que cette optimisation budgétaire pourrait, à terme, nuire à leur agilité et entraîner des effets indésirables, comme des retards sur des projets stratégiques de transformation numérique ou sur le développement des compétences internes. **V.M**

« *Acquérir un malware sur le Darknet reste plus efficace que d'utiliser l'IA générative* »

JOHN SHIER OCCUPE LE POSTE DE FIELD CTO THREAT INTELLIGENCE CHEZ SOPHOS, SOCIÉTÉ BRITANNIQUE DE CYBERDÉFENSE.

En 2024, le nombre de cyberattaques a encore augmenté de 15 % par rapport à l'année précédente, selon l'Anssi. Deux ans après la sortie de ChatGPT, comment l'IA va-t-elle être utilisée par les cybercriminels ? En 2025, va-t-on voir émerger des logiciels malveillants codés entièrement par l'IA ? Réponse avec un expert.

L'informaticien : Depuis la sortie de ChatGPT, la peur est grande de voir les cybercriminels exploiter les possibilités de l'IA générative pour accroître le nombre et l'efficacité des attaques. Jusqu'à présent, cette technologie leur a surtout permis de rédiger des courriels d'hameçonnage plus convaincants dans des langues étrangères. Cela peut-il changer en 2025 ?

John Shier : À mon sens, nous allons rester au stade où l'IA sera principalement utilisée pour faire de l'ingénierie sociale, la technologie n'étant pour l'heure pas suffisamment mûre pour conduire seule des cyberattaques de A à Z. Il faut encore qu'il y ait un humain dans la boucle.

Je ne pense pas non plus que la technologie soit au stade où elle permette à des cybercriminels dotés de connaissances très sommaires en programmation de coder leurs propres logiciels malveillants. En effet, l'IA générative n'est pas infaillible, son code n'est jamais parfait, et il faut donc avoir les connaissances nécessaires pour tout relire, repérer les erreurs et les corriger. Il y a donc toujours une barrière à l'entrée à ce niveau-là, et pour les hackers les moins chevronnés, il reste bien plus simple d'acquérir un logiciel clef en main sur le Darknet. D'autant que ce marché s'est beaucoup développé au cours des dernières années avec l'essor du « *malware as a service* ».

La sortie récente de GPT-5 peut-elle constituer une rupture à cet égard ? Le modèle est-il suffisamment performant pour permettre à des cybercriminels assez doués d'attaquer plus vite et plus fort ?

La grande question est, en effet, de savoir quand nous atteindrons le point de bascule à partir duquel les hackers trouveront plus efficace d'utiliser l'IA générative que de coder à la main. À mon sens, cela ne se fera pas en un jour, mais très progressivement. Un nombre croissant de cybercriminels va continuer d'expérimenter avec cette technologie jusqu'à ce que nous atteignions le seuil de plus de 50 % des logiciels malveillants codés par l'IA, à partir duquel il pourrait y avoir une rapide accélération.

Quant à savoir quand ce seuil sera atteint, c'est pour l'heure impossible de répondre, mais je ne m'attends pas à ce que ce soit le cas en 2025, même si la vitesse à laquelle la technologie progresse peut toujours nous surprendre. Il est d'ailleurs vraisemblable que nous ne nous en rendions pas compte tout de suite : nous prendrons sans doute conscience avec plusieurs mois de retard du fait que tous ces logiciels malveillants que nous croyions codés à la main étaient en réalité le fait d'IAs.

Alors que l'IA générative se démocratise et que la puissance informatique nécessaire pour entraîner ces algorithmes devient plus accessible, est-il probable que l'on voie des groupes de cybercriminels coder leurs propres modèles de fondation spécialisés dans les cyberattaques ?

C'est effectivement une évolution probable. Si les modèles généralistes comme GPT vont demeurer beaucoup trop complexes et coûteux à entraîner, de petits modèles focalisés sur des cas d'usage précis pourraient bientôt séduire les groupes de cyberattaquants les mieux organisés et dotés en ressources. C'est une évolution qui serait particulièrement inquiétante. Pour l'heure, le camp du bien dispose d'un monopole sur les scientifiques des données, ainsi que la création et l'entraînement de ces modèles.

Mais si l'on observe la tendance à l'organisation toujours plus performante du milieu du cybercrime au cours des dernières années, avec notamment l'émergence d'un marché du « *malware as a service* » sur le Darknet, que j'évoquais précédemment, il y a peu de chances pour que cela dure. La tendance dans le monde du cybercrime est à la spécialisation de certains groupes sur une tâche bien précise, et le fait d'entraîner des LLMs pourrait devenir l'une d'entre elles.

En 2023, nous avions conduit une étude pour sonder l'attitude des cybercriminels vis-à-vis de l'IA, et le consensus au sein de leur communauté était que la technologie n'était pas suffisamment mûre pour répondre à leurs besoins. En 2024, notre expert qui avait conduit cette étude a poursuivi ses recherches et constaté un intérêt croissant pour l'IA sur les forums de cybercriminels. Le consensus reste que la technologie doit encore progresser, mais ils se familiarisent avec elle et sont prêts à la mobiliser dès qu'ils sentiront le moment venu. Il faut donc conserver une grande vigilance. **G.R**

Les Cyber Cat'Bonds

pour assurer tous les risques cyber

L'inflation des risques cyber impose aux assureurs de trouver davantage de capitaux (ou capacités) pour les couvrir. Beaucoup plaident pour rééditer la réussite des Cat'Bonds ou Obligations Catastrophes imaginées pour financer une partie de ces risques. Mais cette réplication ne va pas de soi.

La menace d'une crise cyber systémique, pire que Crowdstrike, inquiète aussi bien les assureurs que leurs réassureurs. L'été dernier, cette faille de sécurité chez l'éditeur américain a entraîné des sinistres estimés à 5,4 milliards de dollars. *« Et encore, cela ne prend pas en compte les dommages et intérêts que devront verser à leurs clients des banques ou des compagnies aériennes victimes de ce 'pépin' »*, souligne Patrick Vajda, un vétéran du courtage d'assurance et des risques exotiques tels les grands évènements sportifs internationaux.

Compléter les capacités de l'assurance

Selon, le Forum de Davos, les catastrophes numériques systémiques, avec leurs dommages par ricochet, amputeraient le PIB mondial de 9 500 milliards de dollars à la fin de l'année, de 17 000 milliards en 2030. *« Il faut tempérer ces chiffres, parce que comme pour les catastrophes naturelles, seule une partie des dégâts est assurée, donc indemnisable »*, nuance Philippe Cotelle, administrateur de l'AMRAE, l'association des risks managers français, et co-président de sa commission (cf. interview). Les solutions classiques d'augmentation des primes, bien plus hautes que pour d'autres activités, montrent très vite leurs limites dans un marché hyperconcurrentiel. La création de captives comme Miris, à l'initiative de douze grands groupes européens, comme Airbus ou BASF, qui y logent et y mutualisent une partie de leurs risques numériques, ne peut absorber qu'une partie de ces risques.

D'où le lancement des premiers « Cyber Cat'Bonds ». Entre janvier 2023 et septembre dernier, dix de ces émissions obligataires ILS (Insurance Linked Securities), ont permis à des assureurs et réassureurs de 'transférer' provisoirement près de six cents millions de dollars de risques cybers vers les marchés financiers. Inspirés des Cat'Bonds ou Obligations Catastrophes Naturelles, dans lesquels sont logés ou titrisés temporairement des risques séismes, inondations, tempêtes, ces titres permettraient de compléter les capacités de l'assurance tout en réduisant leur exposition à ces risques et en améliorant leur bilan. Ce qui n'a pas échappé à la Geneva Association, qui réunit les plus grands assureurs et réassureurs européens. Dans un rapport publié à l'automne dernier, son directeur général, Jad Ariss, l'affirme sans ambages : *« la croissance de ce marché impose de nouvelles sources de financement. Des outils comme les Cat'Bonds pourraient augmenter significativement les capacités pour absorber ces risques cybers et contribuer à combler leur énorme déficit d'assurance ».*

Ne pas « répliquer des Cat'Bonds aux risques cybers »

A condition toutefois de lever quelques inconnues. Car les catastrophes naturelles n'ont pas grand-chose à voir avec les risques cybers. *« Les premières sont des risques, le cyber est essentiellement une menace,* insiste Charles de La Horie, le DG de Miris. *Pour les Cat'Nats, la question est de savoir si elles se produiront ou pas. Au contraire des seconds : il faut se demander quand ils se produiront, avec quelle fréquence et surtout avec quelle intensité ».* Mais voilà, es marchés adoptent les Cat'Bonds parce que les évènements qu'ils couvrent sont clairement définis, chiffrés et prévisibles. Rien de tel avec du cyber systémique. La Geneva Association a également sondé de nombreux investisseurs institutionnels avant de rendre leur rapport. Leurs conclusions sont sans appel : ils n'accepteront de prendre en charge *« que des risques extrêmes mais très rares ».* Voire presque impossibles ! Ce qui explique, in fine, l'écart considérable entre Cat'Bonds et Cyber Cat'Bonds : les risques transférés par les seconds pèsent moins de 1 % des quelque 42 milliards de dollars levés par les premiers l'an dernier…

Pour que l'embryon de marché des Cyber Cat' Bonds monte en puissance, il ne faut donc surtout pas *« répliquer des Cat'Bonds aux risques cyber »,* martèle la Geneva Association. *« Il s'agit avant tout de titrisation comme avec les subprimes »,* rappelle Olivier Muraire, l'ancien directeur France et Europe du sud de Liberty Mutual. Avant de rappeler quelques traits des ILS : *« L'investisseur, mais aussi l'émetteur, assureur ou réassureur, doit parfaitement savoir ce qui y est logé : quels sont les risques, quels sont les types d'assurés, quels sont les montants de pertes qui seront supportés, etc. Ce qui suppose de disposer d'outils parfaitement fiables et précis pour bien quantifier ces risques ».*

Mais voilà, ces notions de risques, mais aussi d'assurances, divergent profondément d'un pays à un autre, d'une compagnie à une autre. Autant de points qui freinent aussi le développement de l'assurance des risques cybers. Le directeur des risques cyber de la Geneva Association et auteur de ce rapport, Darren Pain, le reconnaît : *« l'assurance doit mener un vaste chantier de clarification et d'harmonisation de ces termes, de son lexique, des clauses de ses polices ».*

Une construction long terme des Cyber Cat'Bonds

A quoi s'ajoute que : *« ce sont des instruments financiers. Ils permettent à l'assurance d'emprunter des capitaux en échange d'un loyer. Ce qui augmente le coût de ce financement »,* rappelle Patrick Vajda, souligant que *« les primes à verser peuvent rendre l'opération dissuasive, même pour des risques maîtrisés ».* La brusque remontée des taux directeurs après l'invasion de l'Ukraine a considérablement tendu les marchés obligataires. Et incité certains émetteurs à se tourner vers des financements plus conventionnels.

Reste un dernier verrou : celui de la liquidité de ces titres ILS, en particulier sur leur marché secondaire. Lorsqu'elles sont émises ou lancées, ces obligations trouvent rapidement preneurs. Ensuite, le flux des échanges se réduit considérablement. Au point de constituer un nouveau risque lorsque le stress grandit dans les salles de marché. Voire d'amplifier une crise naissante. La liquidité secondaire des Cat'Bonds s'est améliorée tout doucement depuis leur apparition au début des années 90. La légitimité des Cyber Cat'Bonds se construira elle aussi sur le long terme. Sans brûler les étapes et en leur donnant les atouts qui les distinguent des Cat'Bonds sans les opposer. **V.B**

« Des outils pour les assureurs et réassureurs, pas pour les assurés »

Philippe Cotelle, président de la commission cyber risques de l'AMRAE

Les assureurs disposent-ils d'assez de capacités pour couvrir les risques cyber ?

Oui. Le marché peut complètement absorber les demandes de capacités, car peu d'entreprises ont souscrit des contrats individuels nécessitant de 500 à 600 millions de dollars de capacités. Ces capitaux se trouvent facilement à Londres, Francfort, Paris, etc.

Et pour les risques cyber systémiques ?

Non. Si toutes les victimes d'un éventuel sinistre comme Crowdstrike subissent de lourds dommages simultanés à indemniser, alors elles sont insuffisantes.

De futurs Cyber Cat'Bonds pourraient-ils combler ce déficit ?

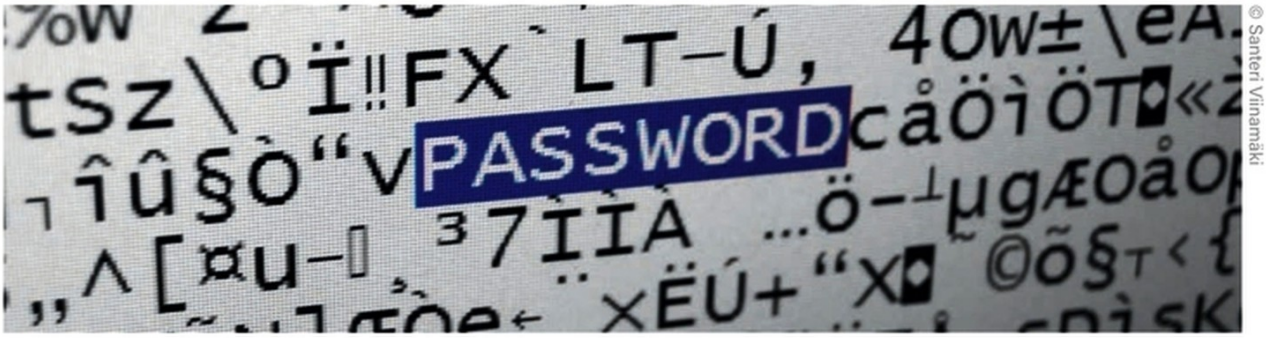
Ils existent déjà. Mais ces outils répondent aux besoins des assureurs et des

réassureurs, pas à ceux des assurés. C'est ce qui les différencie des Cat'Bonds : on trouve même des Cat'Bonds pour couvrir les pylônes d'un opérateur télécom ou d'un électricien en cas de tornades, mais pas de cyber Cat'Bonds pour les risques d'une compagnie aérienne ou d'une banque. Du coup, une simple réplique des Cat'Bonds pour les cyber-risques montrerait vite ses limites.

Que voulez-vous dire ?

On ne peut pas établir un parallèle entre les catastrophes naturelles et les risques cyber : les premières sont localisées géographiquement, et occasionnent des dommages immédiats, avec, en face, des capacités clairement identifiables et réutilisables plusieurs fois. Les seconds ignorent les frontières, causent aussi des dommages par ricochet et consomment immédiatement toutes leurs capacités. Il faudra donc adapter les Cat'Bonds aux spécificités des risques cyber. Cela imposera un examen très approfondi des risques et des assurés logés dans ces titres, afin d'éviter une réédition de la crise des subprimes. Il faudra aussi évangéliser les investisseurs non seulement aux risques cyber mais aussi à l'assurance, à son fonctionnement, à ses spécificités pour éviter toute imprudence.

Résurgence des **attaques par force brute**



© Santeri Viinamäki

Bien que rudimentaires en apparence, les attaques par force brute demeurent un outil de choix pour les cybercriminels. En constante évolution, cette menace qui consiste à tester toutes les combinaisons possibles de mots de passe ou de clés de sécurité jusqu'à trouver la bonne peut causer des dégâts considérables dans les entreprises. Celles-ci doivent mettre en place différentes stratégies de défense pour atténuer leur impact.

Dans le domaine de la cybersécurité, les attaques par force brute rappellent que la simplicité peut être une arme redoutable. Les cybercriminels exploitent la puissance des GPU capable d'exécuter des milliers d'opérations simultanément pour cibler les failles Remote Desktop Protocol (RDP) par le biais d'attaques par force brute. Christophe Gaultier, directeur d'OpenText Cybersecurity France & Belux, tire la sonnette d'alarme sur les dangers liés à ces cyberattaques et leur sophistication croissante : *« Une attaque par force brute est une cyberattaque par laquelle l'attaquant tente d'obtenir un accès non autorisé à un système ou à des données, en essayant systématiquement toutes les combinaisons possibles de mots de passe ou de clés. »* Cette méthode repose sur le simple pouvoir de répétition et sur la capacité de calcul permettant d'essayer des milliards de combinaisons dans un court laps de temps. *« Cela revient à essayer toutes les clés d'un trousseau jusqu'à ce que l'on trouve celle qui ouvre la porte. »*

Une attaque, trois méthodes

« Les cybercriminels exploitent des techniques toujours plus perfectionnées pour réaliser ces attaques, ce qui en fait un véritable défi pour les organisations », souligne Christophe Gaultier. Pour mener ce type d'attaques, les cybercriminels utilisent trois variantes. À commencer par « l'attaque simple », qui consiste à réaliser des essais successifs de combinaisons aléatoires de caractères jusqu'à trouver la bonne. Plus raffinée, la seconde, « attaque par dictionnaire », utilise une liste de mots de passe préexistants, de phrases ou de combinaisons qui sont souvent issues de fuites de données. La troisième méthode, appelée « attaque hybride », combine les deux précédentes approches en modifiant légèrement les mots de passe courants pour essayer de deviner les mots de passe plus complexes.

Les GPU au service des attaques par force brute

Réputées pour leur puissance, notamment dans le développement des technologies d'intelligence artificielle, les unités de traitement graphique (GPU) constituent de redoutables outils pour mener des attaques par force brute. *« Les GPU ont révolutionné non seulement les jeux et la conception graphique, mais aussi le monde de la cybersécurité. Leurs puissantes capacités de traitement parallèle les rendent particulièrement aptes à répondre aux exigences de calcul des attaques par force brute. Contrairement aux unités centrales de traitement (CPU) qui traitent les tâches de manière séquentielle, les GPU peuvent effectuer des milliers d'opérations simultanément, ce qui réduit considérablement le temps nécessaire pour déchiffrer les mots de passe ou les clés de chiffrement »*, ajoute le directeur d'OpenText Cybersecurity. Ce procédé représente une menace importante pour les systèmes protégés par des mots de passe faibles ou couramment utilisés, et souligne la nécessité de mettre en place des politiques de mots de passe robustes et des mesures de sécurité avancées, telles que l'authentification multifactorielle (MFA) et des méthodes de chiffrements résistantes aux attaques menées par les GPU.

Les failles du protocole RDP dans le viseur des hackers

Le protocole propriétaire RDP développé par Microsoft qui permet à un utilisateur de se connecter à un autre ordinateur via une connexion réseau et une interface graphique est la cible privilégiée par les attaquants. Pour Christophe Gaultier, il représente une porte d'entrée particulièrement dangereuse pour les attaques par force brute : *« Si le RDP est un outil puissant pour l'administration et l'assistance à distance, il est également devenu un vecteur privilégié pour les attaques par force brute, car il est couramment utilisé dans les entreprises pour permettre le travail et l'administration de systèmes à distance. Par ailleurs, il exige généralement que le port 3389 soit ouvert, ce qui en fait un point d'entrée visible pour les attaquants à la recherche de vulnérabilités. »* Enfin, la violation d'une session RDP peut donner aux attaquants un contrôle direct sur l'ordinateur d'une victime, ce qui permet de déployer des logiciels malveillants, des ransomwares ou de voler des informations sensibles.

Conséquences financières et réputationnelles

Les implications des attaques par force brute vont bien au-delà du vol direct de fonds. Elles engendrent des pertes financières importantes mais pas que. *« Les implications financières des attaques par force brute peuvent être considérables, allant du vol financier direct à une atteinte substantielle à la réputation entraînant une perte d'activité. Dans certains cas, les attaquants cherchent à obtenir un accès non autorisé aux systèmes financiers ou aux plateformes de paiement. En craquant les identifiants de connexion par la force brute, ils peuvent transférer des fonds, manipuler des transactions ou voler des informations financières sensibles, ce qui entraîne des pertes financières directes. »*, alerte l'expert.

Pour Christophe Gaultier, Directeur Opentext Cybersecurity, les attaques par force brute constituent encore et toujours une tactique de choix dans le paysage actuel de la cybersécurité. Pour que les entreprises puissent s'en prémunir, il recommande d'en connaitre les mécanismes et la méthodologie.

Une protection multiniveau

Face à cette menace persistante, Christophe Gaultier recommande une approche de protection à plusieurs facettes : *« Tout d'abord, il est essentiel d'imposer l'utilisation de mots de passe forts, complexes et uniques, tout en privilégiant l'authentification multifactorielle (AMF) pour renforcer les barrières d'accès. Ensuite, la mise en place d'une politique de verrouillage des comptes, qui désactive temporairement un compte après plusieurs tentatives infructueuses, limite considérablement les opportunités pour les attaquants. »* L'authentification au niveau du réseau (NLA) constitue également un rempart important, exigeant que les utilisateurs soient authentifiés avant de pouvoir accéder à une session RDP. Par ailleurs, l'utilisation d'un VPN pour restreindre l'accès RDP au seul trafic sécurisé réduit la visibilité de ce protocole sur Internet. Enfin, des outils de surveillance et d'alerte permettent de détecter les tentatives répétées de connexion et de notifier rapidement les administrateurs, leur donnant ainsi les moyens de répondre aux menaces en temps réel. Une fois combinées, ces mesures forment un bouclier essentiel contre les attaques de force brute.

J.C

Sécurité des mots de passe : adopter des solutions modernes pour limiter les risques

Les attaques par force brute exploitent une faiblesse universelle : des mots de passe trop simples et des politiques de sécurité mal adaptées. Pour contrer ces menaces, l'une des stratégies les plus efficaces consiste à adopter des mots de passe longs (15 à 20 caractères), tout en abandonnant les directives obsolètes, comme l'obligation d'inclure des majuscules, des chiffres ou des caractères spéciaux. De même, les politiques de rotation fréquente des mots de passe, souvent contre-productives, devraient être remplacées par des approches plus modernes. Cependant, ces recommandations se heurtent à une difficulté majeure : les employés ont souvent tendance à choisir des mots de passe simples, prévisibles ou réutilisés sur plusieurs comptes. L'utilisation d'un gestionnaire de mots de passe (Dashlane, LastPass Business...) simplifie considérablement la mise en place d'une stratégie de mots de passe robuste et efficace. Ces outils permettent de générer des mots de passe uniques, longs et complexes, tout en les stockant de manière sécurisée. De plus, certains gestionnaires offrent des fonctionnalités avancées, comme la détection proactive des risques liés aux identifiants, le contrôle des accès, et le partage sécurisé des informations d'authentification. Ils réduisent significativement les risques tout en simplifiant la gestion des mots de passe au quotidien.

La protection des données personnelles de plus en plus un sujet de société en France

Par Paul-Olivier Gibert, président de l' AFCDP

La question de la protection des données personnelles s'impose de plus en plus comme un sujet de société incontournable en France et en Europe. Avec l'explosion du numérique, nos vies quotidiennes se déploient sur des plateformes qui collectent, analysent et utilisent nos données à des fins variées. Si ces outils offrent des services pratiques et innovants, ils suscitent aussi une prise de conscience sur les risques liés à l'utilisation de nos données personnelles.

L'émergence de la protection des données comme sujet sociétal trouve ses racines dans des avancées législatives. En Europe, le RGPD, entré en vigueur en 2018, impose aux entreprises des obligations plus rigoureuses que celles qui prévalaient en France, depuis 1978, avec la Loi Informatique et Libertés, et dans l'Union européenne depuis la directive de 1995. En France, la CNIL veille à la mise en œuvre de ce cadre et à la protection des droits des citoyens, tels que le droit d'opposition, à l'effacement ou à la portabilité des données. Ces avancées juridiques renforcent la transparence et responsabilisent les acteurs économiques dans leur manière de collecter et traiter les données. D'autre part, la charge de la preuve pesant sur les organismes collecteurs, rend plus facile le dépôt de plainte par les citoyens sur le site de la CNIL.
Le respect de ces règles ne se fait pas toujours sans heurts. La CNIL a multiplié les sanctions ces dernières années. En 2020, Google et Amazon ont été condamnés à des amendes record de 100 et 35 M€. En 2022, c'était au tour de Meta d'écoper d'une amende de 60 M€. Ces affaires mettent en lumière les pratiques abusives de certaines entreprises, et sensibilisent le grand public aux enjeux de la protection des données.

Une sensibilisation croissante des citoyens

Les Français sont de plus en plus conscients des risques liés à la collecte massive de leurs données personnelles. Des scandales comme celui de Cambridge Analytica en 2018 ont joué un rôle de révélateur. Aujourd'hui, près de 81 % des citoyens disent être préoccupés par le sort réservé à leurs données personnelles, selon un sondage IFOP de 2022.
En parallèle, la montée en puissance des cyberattaques et l'augmentation du nombre de fuites de données accentuent les inquiétudes. Les hôpitaux, administrations publiques et entreprises françaises sont devenus des cibles privilégiées des hackers. Ces incidents révèlent la vulnérabilité des systèmes et mettent en péril des données sensibles, comme celles des patients. En réponse, les organismes, publics et privés, investissent de plus en plus dans des solutions de cybersécurité.
La protection des données ne se limite pas à des enjeux technologiques ou juridiques. Elle devient aussi un sujet éducatif. La CNIL s'investit dans le conseil, la prévention et la sensibilisation. Les programmes scolaires incluent désormais des modules de sensibilisation à l'usage responsable d'Internet et à la gestion des données personnelles. Les débats publics se multiplient, qu'il s'agisse des controverses sur la reconnaissance faciale ou des outils de surveillance. Ces discussions renforcent la place de la protection des données dans l'espace médiatique.

Une responsabilité économique et sociétale

Les entreprises, elles aussi, prennent conscience de leur responsabilité. La mise en conformité avec le RGPD, au-delà d'affecter la valeur comptable de l'entreprise devient un argument commercial. Des certifications comme celles portant sur la formation, ou sur les compétences des délégués à la protection des données, garantissent une démarche respectueuse des données. Cette évolution répond à une demande croissante des consommateurs pour plus d'éthique et de transparence. Ce contexte donne tout son sens à la fonction de délégué à la protection des données, à la croisée entre la protection des intérêts de l'organisme qui l'a désigné, et la protection des droits et de la vie privée des salariés, collaborateurs, clients ou prospects.

La protection des données est bien plus qu'un enjeu technique ou réglementaire. Elle traduit une évolution de nos valeurs, avec un souci accru pour la vie privée et les libertés individuelles. Si des efforts restent à faire, notamment dans l'éducation et la cybersécurité, la France progresse vers une culture de la protection des données.

Donald Trump veut une IA sans entrave

Le 47è président des États-Unis, Donald Trump, a abrogé, lundi 20 janvier 2024, un décret signé par Joe Biden sur la sécurité de l'intelligence artificielle. Il a, en outre, révoqué certaines politiques et directives faisant « obstacle à l'innovation américaine » dans ce domaine. Des décisions qui préfigurent un développement de l'IA davantage tourné vers une compétitivité sans entrave, au détriment de la sécurité.

À peine arrivé à Washington, Donald Trump veut faire table rase du passé, et plus particulièrement de celui de Joe Biden. Le jour même de son investiture, le nouveau locataire de la Maison-Blanche a abrogé de nombreux décrets pris par son prédécesseur. Tous ont été recensés dans un communiqué intitulé : *« Premières annulations de décrets et d'actions préjudiciables »*. Pour rappel, le décret présidentiel aux États-Unis a force de loi et permet au président d'orienter directement l'action publique en donnant notamment des directives aux fonctionnaires et agences fédérales. En revanche, ces Executive Orders peuvent être annulés par le Congrès ou un tribunal, si jugés non conformes à la Constitution.

Une première tentative de réglementation

Parmi les textes jetés aux orties : l'Executive Order 14110 du 30 octobre 2023 sur le développement et l'utilisation sécurisée et fiable de l'intelligence artificielle.
Ce texte posait les ébauches d'une feuille de route pour développer une IA sûre et responsable et en prévenir les dérives. Parmi les principales mesures : l'obligation pour les développeurs de tester et d'évaluer la sécurité des systèmes d'IA pour en atténuer les risques (via des évaluations red-teaming, notamment) et l'instauration de mécanismes de transparence tels que l'étiquetage des contenus générés. Avec ce décret, l'administration Biden voulait faire en sorte que les lois fédérales en matière de droits civiques soient respectées, et éviter ainsi les biais discriminatoires, renforcer la sécurité des données, limiter leur collecte et leur utilisation abusive. Le texte missionnait également le NIST (National Institute of Standards and Technology) pour définir des normes, et chargeait d'autres agences fédérales d'évaluer les risques biologiques, radiologiques, nucléaires, de cybersécurité ou d'infrastructure critique que l'IA pourrait poser.

USA vs UE : une opposition de modèle

Soucieux de préserver avant tout le leadership des États-Unis dans ce domaine en pleine expansion, Donald Trump a enfoncé encore un peu plus le clou, jeudi 23 janvier 2025, en signant un décret qui révoque *« les politiques et directives existantes qui agissent comme des obstacles à l'innovation américaine en IA »*. Et ce, afin d'*« améliorer la domination mondiale de l'Amérique en la matière [et] de promouvoir l'épanouissement humain, la compétitivité économique et la sécurité nationale »*. Les plus proches conseillers du président sur les questions des nouvelles technologies devront examiner toutes politique, directive, ou règlement et autres qui auraient été entrepris en vertu du décret du 30 octobre 2023 et feraient obstacle à la compétitivité des États-Unis. A leur charge de les suspendre, les réviser ou les abroger quand cela sera possible.
Les quelques efforts concédés par l'administration Biden pour exploiter l'IA de manière responsable devaient se faire en collaboration avec les alliés des États-Unis, afin de créer des cadres communs de gouvernance. L'annulation de ce décret creuse un peu plus le gouffre entre le modèle de développement adopté par l'oncle Sam et celui de l'Union européenne qui tente d'encadrer cette technologie. L'AI Act, dont les premières exigences s'appliqueront dès le 2 février, a pour ambition d'établir des règles en matière de sécurité des usages des systèmes d'IA, tout en limitant l'impact sur l'innovation dans le domaine. Une critique qui lui est souvent faite.

V.M

Avec Dora, la gestion contractuelle au centre

Entré en vigueur le 17 janvier 2024, le Digital Operational Resilience Act (Dora) a revu les accords contractuels conclus entre les entités financières et leurs prestataires tiers de services TIC. Et impose toute une série de clauses afin de garantir une meilleure gestion et la maîtrise des risques.

Julie Jacob, avocate spécialisée en droit du numérique

La généralisation des technologies de l'information et de la communication (TIC) a eu pour effet pervers d'exposer de plus en plus les entités financières qui les utilisent à des risques cyber. Pour tenter d'y remédier, l'Union européenne a fait ce qu'elle sait faire de mieux : réglementer.

Le règlement Dora introduit des exigences spécifiques au secteur financier, afin de renforcer leur cybersécurité et leur résilience opérationnelle face aux risques liés à l'utilisation des TIC. Entré en vigueur le 17 janvier 2025, il impose toute une série de règles à respecter concernant la gestion du risque informatique, le reporting des incidents, les tests de résilience et la gestion du risque tiers porté par les prestataires de services informatiques.

Au-delà des prescriptions techniques, le texte fixe des exigences contractuelles entre les entités financières et leurs prestataires TIC. *« Dora donne des outils pour maîtriser les risques. Parmi ceux-ci, il introduit des clauses afin de revoir la conformité des contrats existants (et futurs) avec les prestataires informatiques »*, développe Julie Jacob, avocate spécialisée en droit du numérique et des nouvelles technologies, et fondatrice du cabinet Jacob Avocats. Dans le détail, les contrats doivent désormais décrire les services fournis, préciser si des fonctions critiques sont sous-traitées et sous quelles conditions. Ils doivent indiquer où les services seront fournis et les données traitées — avec une obligation de notification en cas de changement —, garantir la sécurité, la confidentialité, l'accès et la récupération des données, même si le prestataire cesse ses activités. Les TIC ont aussi pour obligation d'indiquer les niveaux de service, de fournir une assistance en cas d'incident, de coopérer avec les autorités et l'entité financière, et de renseigner les droits de résiliation.

Des clauses au goût du risque

Les prestataires qui fournissent des prestations critiques — lesquelles ont un impact majeur sur le secteur — sont soumis à une supervision renforcée des autorités européennes, telles que les Autorités européennes de surveillance (les AES), et à des exigences contractuelles spécifiques. Les clauses devront ainsi renseigner des niveaux de services détaillés avec mise à jour, les éventuelles révisions et objectifs, les délais de préavis et de notification à l'entité financière, une notification en cas d'impact sur les capacités à fournir les services TIC, l'obligation de mise en place de plans d'urgence et de mesures de sécurité appropriées. Le contrat inclut, en outre, le droit de suivi des performances du prestataire tiers de services TIC (audit inclus) et prévoit, en cas de perturbation pour l'entité financière, des stratégies de sortie et les conditions d'une migration vers un autre prestataire.

Les entités financières, elles, sont responsables des prestataires qu'elles choisissent. C'est pourquoi elles doivent mener des audits pour s'assurer qu'ils respectent des normes de sécurité élevées. Pour se couvrir, elles peuvent aussi intégrer des clauses de résiliation dans les contrats, par exemple. *« Le règlement Dora permet une résiliation simplifiée des contrats, si des modifications importantes non validées surviennent, comme un changement de sous-traitant non conforme par le prestataire de services TIC »*, détaille Julie Jacob, ou encore en cas de constatation d'une gestion insuffisante des risques ou d'un manque de coopération.

Julie Jacob précise : *« Si les entités financières sont sensibilisées au sujet de Dora, les prestataires de TIC, qui ont parfois des clients dans plusieurs secteurs, ont moins cette culture. C'est à eux, surtout, de prendre conscience que leurs contrats vont devoir être mis à jour. »*

V.M

Les **cryptomonnaies** sont-elles menacées par l'arrivée du quantique ?

Avec l'arrivée de l'informatique quantique d'ici quelques années, que va-t-il se passer pour la blockchain et les cryptomonnaies ? Leur fiabilité et leur existence mêmes sont fondées sur des algorithmes de cryptage réputés inviolables. Seront-elles annihilées brutalement ou résisteront-elles à ce tsunami en matière de sécurité ?

Dans le monde de la cryptomonnaie, une menace a toujours existé et reste omniprésente : la prise de contrôle des validateurs d'une monnaie permet de la hacker et de détruire son écosystème en un battement de cils. Si cela était considéré comme presque impossible jusqu'ici, au vu de la difficulté pour résoudre le hash d'un bloc de bitcoin ou d'autres monnaies, les choses pourraient changer avec l'informatique quantique. Par rapport à un ordinateur classique, un ordinateur quantique est censé pouvoir effectuer un nombre d'opérations bien plus important dans le même laps de temps. Quand un « simple » ordinateur, même le plus puissant, aurait besoin de plusieurs dizaines d'années pour trouver la solution permettant de résoudre un bloc, ou des siècles pour essayer des clés privées en boucle, un ordinateur quantique pourrait faire le même travail en quelques jours ou heures, voire encore moins.

La menace du quantique

La menace réside principalement dans la possibilité de casser la cryptographie des clés publiques. La force de ce système repose sur un ensemble de problèmes mathématiques (des algorithmes de cryptage) que les ordinateurs classiques ne peuvent résoudre rapidement. Parmi les cryptages employés, on retrouve par exemple le cryptage RSA et le cryptage à courbe elliptique (ECC). Le premier, RSA, repose sur la difficulté de factoriser de grands nombres composés. Le décryptage de longues clefs prendrait à l'heure actuelle plusieurs milliers d'années de calculs. Le deuxième, ECC, est employé par le bitcoin, l'ethereum et de nombreuses autres cryptomonnaies. Il repose sur des problèmes logarithmiques demandant là aussi une très grande puissance de calcul. Les chercheurs pensent que les ordinateurs quantiques devraient réussir à casser la plupart de ces algorithmes de cryptage dans 10 ou 20 ans. Le plus puissant ordinateur quantique existant est de 156 qubits. Dans 10 à 20 ans, des ordinateurs de plusieurs centaines de milliers, voire de plusieurs millions de qubits, devraient voir le jour.

Créer des blockchains « quantum-resistant »

Certaines blockchains sont déjà résistantes aux algorithmes quantiques. Leurs protocoles reposent sur l'utilisation d'algorithmes judicieusement qualifiés de « post-quantiques ». Elles ont été conçues pour résister à la future puissance des ordinateurs quantiques. Les méthodes de cryptographie auxquelles elles ont recours sont énumérées ci-après. La cryptographie basée sur des treillis peut être représentée par une grille en trois dimensions composée de milliards de nœuds. Le chemin le plus court pour relier deux points devient la clef pour résoudre l'énigme. Cette méthode est tellement complexe que même un ordinateur quantique aura normalement besoin de beaucoup de temps pour réussir à calculer le dit chemin. La cryptographie à base de hachage fonctionne un peu comme une empreinte digitale numérique. Le hachage n'est possible que dans un sens. Comme il n'est pas réversible, il est impossible de

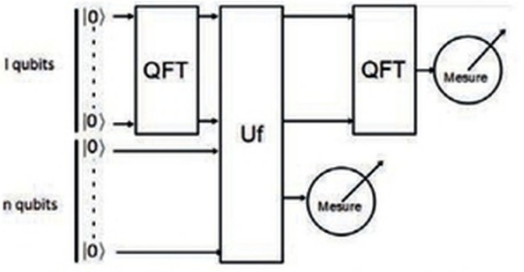


FIGURE 1 – L'algorithme quantique de recherche de période est l'algorithme

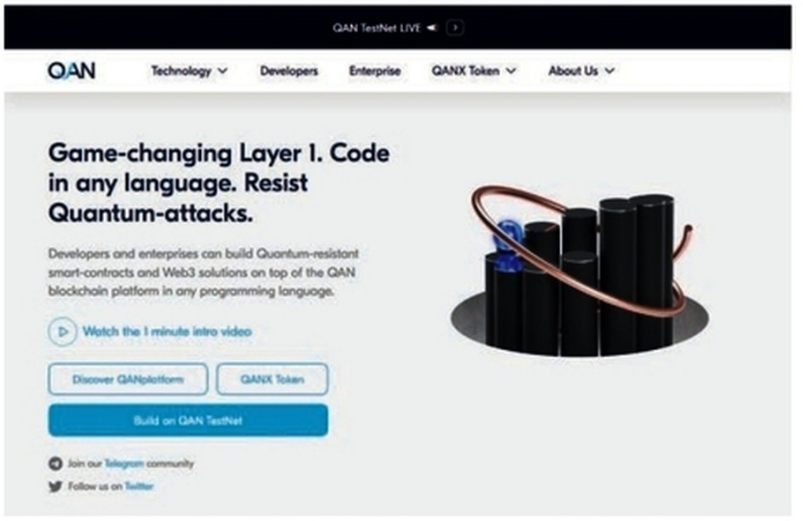
$$\frac{1}{\sqrt{q}}\sum_{a=0}^{q-1}|a\rangle\otimes|0^n\rangle \tag{5}$$

Plus intuitivement, la première transformée de Fourier permet de faire en sorte que le premier registre balaie tout les entiers de 0 à $q-1$.

Maintenant, appliquons U_f à l'ensemble des deux registres. On obtient immédiatement :

$$\frac{1}{\sqrt{q}}\sum_{a=0}^{q-1}(|a\rangle\otimes|f(a)\rangle) \tag{6}$$

L'algorithme de Shor a été conçu dans le but de casser les systèmes de cryptographie fondés sur la factorisation de grands nombres comme RSA.



Les développeurs peuvent créer des smart-contracts résistants au quantique et des solutions Web3 sur la plateforme de blockchain QAN avec n'importe quel langage de programmation

revenir au mot de passe à partir du hachage. Cette méthode n'est pas vraiment nouvelle, mais elle a été complexifiée pour la rendre plus résistante face au quantique. La cryptographie basée sur un code, sur une méthode incluant la dissimulation d'un message dans un signal dit « bruyant » nécessite de posséder la clef pour « entendre » le message en question. Ce système n'est, lui non plus, pas vraiment nouveau. Il existe depuis une quarantaine d'années. Néanmoins, la taille qui lui est nécessaire le rend suffisamment complexe pour lui permettre de résister au quantique. La cryptographie polynomiale multivariée est, quant à elle, une espèce de puzzle d'équations complexes non linéaires et à plusieurs variables. Là encore, même un ordinateur quantique aura bien du mal à résoudre des calculs de ce type simultanément.

Ces quatre techniques de cryptage mériteraient des explications plus détaillées au vu de leur complexité, mais nous manquerions de place pour le faire dans cet article et vous demanderons d'accepter ce résumé (très) rapide : elles reposent toutes sur l'utilisation de méthodes que l'augmentation considérable du nombre d'opérations par secondes ne permet pas de résoudre plus rapidement. Elles devraient donc de facto être résistantes au quantique. Les jetons de cryptage actuels et futurs devront donc être « quantum-résistants », afin de préserver la sécurité, la viabilité et l'intégrité des réseaux informatiques et des cryptomonnaies. C'est l'intégralité de la blockchain qui devra résister. Les transactions déjà archivées sont bien évidemment également concernées. Les futures attaques quantiques pourraient compromettre des transactions en cours de validation en les falsifiant, mais aussi des transactions déjà validées. L'immuabilité étant à la base du fonctionnement des registres partagés, comme par exemple pour des actes notariaux, c'est toute l'infrastructure de la blockchain qui pourrait s'écrouler et devenir inutile.

Les défis liés aux jetons quantum-résistants

Malgré cette résistance, plusieurs complications sont à signaler, impliquant de devoir encore améliorer certaines techniques. Les algorithmes post quantiques, tels que le treillis ou le codage, requièrent l'emploi de ressources très importantes pour être opérationnels. Cela nécessitera donc l'utilisation de machines très puissantes de tous les côtés pour ne pas trop ralentir leur fonctionnement, et entraînera une forte consommation d'énergie. Il sera donc difficile de les utiliser dans tous les contextes (surtout matériels). La taille des clefs de signatures post quantiques, très grandes, (bien plus que la moyenne actuelle) pose elle aussi un problème. Ces clefs « pèsent » plusieurs kilo-octets et peuvent entraîner des problèmes de stockage et de transmission, ce qui peut les rendre incompatibles avec les systèmes existants qui auraient, eux, plutôt tendance à réduire la taille du moindre paquet de données. Enfin, il n'existe pas encore vraiment de consensus général concernant la résistance aux attaques quantiques. Le NIST (National Institute of Standards and Technology) y travaille sérieusement mais, à ce jour, rien n'a encore été décidé. Les projets existants pourraient prendre des directions séparées, ce qui menacerait fortement l'interopérabilité actuelle et l'efficacité générale contre ces attaques. L'infrastructure actuelle n'ayant pas anticipé ces changements, la transformation des réseaux existants pourrait nécessiter des bouleversements très importants, y compris sur le bitcoin.

En conclusion, les ordinateurs quantiques pourraient facilement casser les algorithmes de cryptographie actuellement utilisés par les blockchains. Pour faire face à cette menace, elles devront être mises à niveau en adoptant d'autres algorithmes résistants cette fois aux attaques quantiques. C'est un travail important à réaliser pour chaque système, mais totalement incontournable pour garantir leur pérennité. ▮

T.T

Les bons élèves

Quelques cryptomonnaies ont pris un peu d'avance et seraient d'ores et déjà prêtes à faire face à cette menace. La Quantum Resistant Ledger utilise des signatures numériques obtenues grâce au hachage. Un timbre numérique garantit l'inviolabilité des transactions. La QANplatform intègre, quant à elle, le système de treillis dans sa blockchain. La IOTA se distingue des autres en employant le schéma de signature WOTS (Winternitz One-Time Signature Scheme) avec Tangle, un schéma de vérification censé garantir l'intégrité de son infrastructure. Néanmoins, tout ceci reste encore purement théorique, et il faudra voir si l'avenir et la pratique confirment leur supposée résistance. Des modifications futures de leur protocole et des algorithmes quantiques pourraient tout changer.

Contre les deepfakes, *« la meilleure défense, c'est de ne pas être une cible facile »*

Philippe Luc, cofondateur d'Anozr Way

Anozr Way fournit des outils pour aider les dirigeants et leurs collaborateurs à contrôler leur empreinte numérique et mieux protéger leurs données qui pourraient être utilisées dans le cadre d'attaques par ingénierie sociale, telles que le phishing, le ransomware et, désormais, le deepfake. Une nouvelle menace qui a toutes les chances de devenir un risque systémique pour les entreprises. Cofondateur d'Anozr Way, Philippe Luc nous en dit plus.

L'Informaticien : Pourquoi est-il important de maîtriser son empreinte numérique dans un contexte de montée en puissance des deepfakes ?

Philippe Luc : Parce que faire de l'OSINT (renseignement d'origine sources ouvertes) est devenu extrêmement facile. N'importe qui peut se renseigner sur quelqu'un, et les nombreux leaks de ces dernières années amplifient encore la disponibilité des données personnelles en ligne. Les cybercriminels vont sur le deep web ou le dark web pour récupérer un maximum d'informations afin de préparer leurs attaques. Le deepfake constitue une nouvelle méthode d'attaque par ingénierie sociale qui consiste à utiliser l'IA pour donner vie à cette ingénierie sociale et la rendre d'autant plus convaincante. Nous avons changé d'échelle dans la capacité à tromper l'humain. En 2023, nous parlions de 500 000 deepfakes vidéos partagés sur les réseaux sociaux. Nous nous attendons à en avoir 8 millions de plus en 2025. Cela peut paraître dérisoire, mais avec la viralité, une seule vidéo peut être vue des centaines de millions de fois.

Concrètement, quels risques font peser les deepfakes sur les entreprises ?

PL : Les dommages sont à la fois financiers et réputationnels. Par exemple, une fausse vidéo d'un dirigeant annonçant la fermeture de plusieurs sites industriels peut entraîner une atteinte financière directe si l'entreprise est cotée en bourse. Et à partir du moment où le deepfake a été publié, même s'il y a un démenti, avec la viralité, le marché aura potentiellement déjà réagi, ce qui entraînera des pertes financières. Le deepfake porte aussi un coup à la réputation du dirigeant, mais également à celle de l'entreprise et de tout son écosystème. Dans ce contexte, cette menace a toutes les chances de devenir un risque systémique pour les entreprises.

Quels sont les grands principes pour maîtriser son empreinte numérique et limiter l'impact des deepfakes ?

PL : La meilleure défense, c'est de ne pas être une cible facile, en laissant le moins possible de matière première facilement exploitable par les pirates, comme des photos, vidéos, et toutes données personnelles qui permettraient d'enrichir le résultat final. Il faut faire en sorte que les deepfakes soient les plus complexes possibles à réaliser. Cela concerne tout le monde, car ce ne sont pas toujours les grands dirigeants qui sont ciblés. Parfois, ce sont des collaborateurs de niveau N-1, dont les informations personnelles sont moins accessibles en milieu professionnel mais davantage dans leur vie privée. Si un cybercriminel scrute leurs réseaux sociaux mal sécurisés et publics, il pourra y trouver des vidéos personnelles, comme un discours donné à l'anniversaire d'un proche, ou des photos qui lui permettront de créer un deepfake voix et image très convaincant. Le discours semblera d'autant plus authentique s'il s'appuie sur des faits réels, comme des événements de votre vie privée ou professionnelle, des projets d'entreprise, le nom de vos enfants, vos centres d'intérêt, autant d'informations potentiellement accessibles publiquement. Si un deepfake intègre ces détails, il aura plus de chances de vous convaincre, même si la qualité de la vidéo est médiocre.

La nécessité pour les entreprises de maîtriser l'empreinte numérique de leurs dirigeants et collaborateurs implique donc un important volet de sensibilisation ?

PL : Il faut travailler avec toutes les équipes en interne. Les comités exécutifs (Comex) doivent être sensibilisés aux deepfakes, savoir comment ils fonctionnent, comment ils sont utilisés et quelle matière première sert à leur création. Ensuite, l'entreprise doit être vigilante quant au type de contenus qu'elle partagera à l'avenir. Un exemple : aujourd'hui, on le voit bien, l'IA a du mal à gérer les doigts ou la synchronisation labiale. Si vous publiez des vidéos de votre patron lors d'une conférence, demandez-lui de mettre son micro devant la bouche. Ainsi, le cybercriminel n'aura pas suffisamment de temps de vidéo de bonne qualité pour produire une bonne synchronisation labiale. Les équipes communication vont donc avoir du travail, tout comme les équipes de sécurité, qui vont devoir se pencher sur des plans de gestion de crise et de sensibilisation en interne.

PURESTORAGE®

Simplifier le stockage des données pour toujours

Le stockage à la demande (STaaS) vous permet de bénéficier d'une flexibilité financière et d'une simplicité opérationnelle pour répondre durablement aujourd'hui et demain, aux besoins de votre entreprise.
Evergreen//One™ associe l'agilité du stockage dans le cloud public à la sécurité et aux performances d'une infrastructure all-flash. Cette solution STaaS offre une véritable expérience de cloud hybride.

www.purestorage.com/fr/products/staas/evergreen/one.html

SMART TECH

DELPHINE SABATTIER

7H30 | 18H30

VOTRE ÉMISSION QUOTIDIENNE DÉDIÉE À L'INNOVATION

Dans l'émission SMART TECH animée par Delphine Sabattier, l'actualité du numérique et de l'innovation prend tout son sens. Chaque jour, des spécialistes décryptent les dernières news, les tendances, et les enjeux soulevés par l'adoption des nouvelles technologies.

N°230 orange™ | N°246 bouygues | N°163 free

B SMART 4. Change