

L'INFORMATICIEN

DevOps

Quelle base de données pour l'IA

Cloud

Google Security Command Center Enterprise

Réseau

LoRA le grand gagnant de l'IoT

DOSSIER

Blockchain Le renouveau

RH

On demande spécialiste Mainframe !

Hardware

La photonique trouve la lumière



STANDARD TÉLÉPHONIQUE | VIDÉO | LIVE CHAT

CONNECTEZ VOTRE ÉQUIPE & VOS CLIENTS



Préparez votre entreprise au **Smart Working**
avec une solution **économique** :

- PBX hébergé ou autogéré
- Visioconférence
- Live Chat sur site web
- Intégration WhatsApp et Facebook
- Solution de centre d'appels
- Applications iOS/Android et web
- Intégration Microsoft 365



OBTENEZ 3CX GRATUITEMENT ! WWW.3CX.FR

3CX

RÉDACTION

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

RÉDACTION : Bertrand Garé (réacteur en chef) et Victor Miget (réacteur en chef adjoint).
avec : Oscar Barthe, Patrick Brebion, Jérôme Cartegini, Michel Chotard, Alain Clapaud, François Cointe et Thierry Thaureau.

SECRÉTAIRE DE RÉDACTION : Boutheïna Saddi

MAQUETTE ET RÉALISATION : Franck Soulier (chef de studio)

PUBLICITÉ

Tél. : +33 (0)1 74 70 16 30 – pub@linformatien.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.linformatien.com/abonnement

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

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

IMPRESSION

Imprimé en France par Imprimerie Chirat (42)
Dépôt légal : 2^{ème} trimestre 2024

Toute reproduction intégrale, ou partielle, faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause, est illicite (article L122-4 du Code de la propriété intellectuelle). Toute copie doit avoir l'accord du Centre français du droit de copie (CFC), 20 rue des Grands-Augustins 75006 Paris. Cette publication peut être exploitée dans le cadre de la formation permanente. Toute utilisation à des fins commerciales de notre contenu éditorial fera l'objet d'une demande préalable auprès du directeur de la publication.

L'INFORMATIEN 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



PRÉSIDENT, DIRECTEUR DE LA PUBLICATION :
Gaël Chervet

Sous le signe de l'innovation

Ce numéro 225 fait la part belle aux technologies de rupture. Si l'intelligence artificielle tient le haut du pavé, d'autres comme la blockchain connaissent un renouveau intéressant au sein des entreprises. Son intérêt et sa présence sont souvent négligés en raison de la spéculation entourant les plateformes publiques d'échange de crypto-monnaies. Nous revenons largement sur son développement actuel dans notre dossier. Ces différentes technologies de rupture ont en commun de s'appuyer sur une pile logicielle nativement Cloud. Récemment s'est tenue la KubeCon, la conférence de la communauté Kubernetes et de sa « maison mère », la CNCF, qui devient l'acteur principal de ce nouvel écosystème centré sur le Cloud.

Une autre technologie, parfois encore éclipsée des radars des observateurs, est la photonique qui continue sa progression, avec des offres de hardware commençant à apparaître sur le marché afin de répondre aux besoins en débit et en latence nécessaires au traitement de vastes volumes de données, notamment dans le domaine de la Gen AI. Il devient donc inévitable d'aborder ce sujet. Shoptalk, le salon mondial de l'industrie du commerce du retail, en est un exemple, avec de nombreuses innovations s'appuyant sur cette technologie pour améliorer le service aux clients.

Vous retrouverez bien sûr tous nos rendez-vous habituels. Mention spéciale pour notre événement Top Tech qui débutera dans les prochains jours. Nous comptons sur vous pour nous proposer de nombreux dossiers que notre jury de DSI analysera afin de déterminer les lauréats de cette nouvelle édition. Nous vous souhaitons une bonne lecture et nous vous invitons à nous faire part de vos suggestions, vos critiques, et peut-être même de vos compliments pour que nous puissions continuer à améliorer notre publication. □

Bertrand Garé
Rédacteur en Chef



SOLUTIONS D'INFRASTRUCTURES RÉSEAUX D'ALCATEL-LUCENT D'ENTERPRISE



Alcatel-Lucent Enterprise, société de droit français forte d'une tradition centenaire d'innovation au service des entreprises. Leader mondial, ALE propose des solutions et services de réseau d'entreprise, communication et cloud souverain.

Accédez à un environnement sécurisé, efficace et de haute performance, nécessaire pour assurer la transformation numérique de votre entreprise.

- Gamme robuste, simplifiée et évolutive
- Fonctionnalités sécurisées orientées métiers
- Facilité de déploiement

CONTACTEZ-NOUS SUR
contact.france@al-enterprise.com
pour rencontrer un expert ALE.
<https://www.al-enterprise.com/fr-fr>

#WhereEverythingConnects

* Le nom et le logo d'Alcatel-Lucent sont des marques commerciales de Nokia utilisées sous licence par ALE.

Alcatel-Lucent
Enterprise

P 36
LORA
LE GRAND
GAGNANT
DE L'IOT

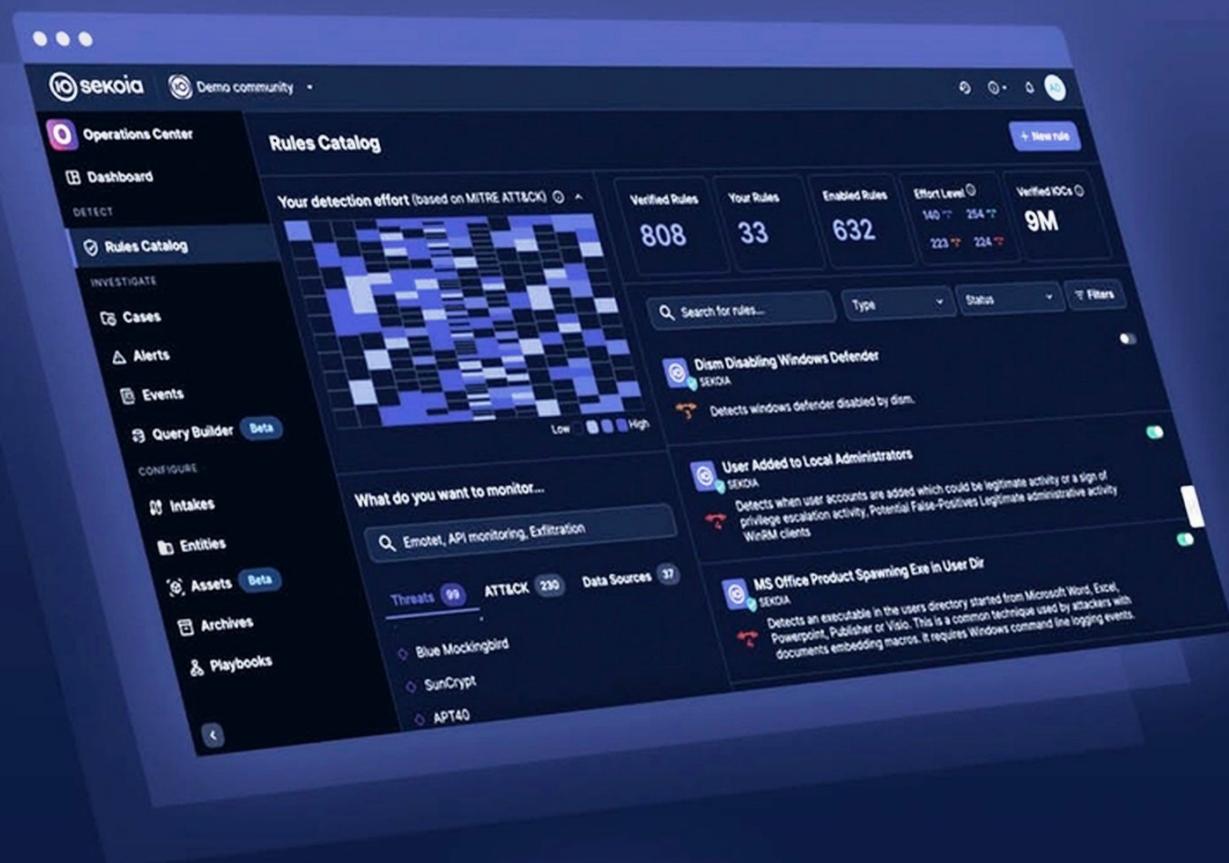


DOSSIER	P 15
Blockchain : le renouveau	
BIZ'IT	P 8
BIZ'IT PARTENARIAT	P 12
TACTIC	P 23
Les jeux sont faits !	
HARDWARE	P 26
Photonique	
Test Philips	
ESN	P 30
Tendances PAC	
Manifeste EY Fabernovel	
RÉSEAU	P 32
Terabit Ethernet	
Haut débit fixe	
Trafic mobile	
LoRa	
LOGICIEL	P 38
TrailBlazer DX	
Adobe	
Workday	
Java 22	

CLOUD	P 46
Kubecon	
Google Cloud	
RETEX	P 52
Ville de Macon	
Arpavie	
DEVOPS	P 56
Base de données IA	
BONNES FEUILLES	P 61
Vous ne me trouverez pas	
sur Amazon	
INNOVATION	P 65
Shoptalk	
Confiance.ai	
Deep Tech BPI	
RH	P 70
Formation Mainframe	
INFOCR	P 75
ABONNEMENTS	P 43



Boostez votre sécurité avec la plateforme SOC Sekoia!



800 règles de détection
prêtes à l'emploi

Contextualisation des alertes
grâce à la CTI native et exclusive

180+ intégrations avec
des technologies tierces

Automatisation facilitée
avec les playbooks intégrés



Sekoia.io a reçu le prix *Customer Value Leadership 2023* décerné par Frost & Sullivan, positionnant la plateforme SOC Sekoia comme le leader incontesté dans l'industrie européenne de la détection et de la réponse étendue (XDR).

LES PROGRÈS DE LA BLOCKCHAIN

AVEC L'ADN, JE SAIS JUSTE QU'IL EST NÉ AVEC LE GÈNE DE LA BÉTISE.

MAIS C'EST TOUT

AVEC LA BLOCKCHAIN J'EN SAIS BEAUCOUP PLUS QUE C'ELA !

JE PEUX CERTIFIER QU'IL A PLUS DE 18 ANS ET QU'IL PEUT ALLER SUR LES SITES PORNOS.

JE PEUX VÉRIFIER QU'IL N'A PAS FALSIFIÉ SES DIPLOMES NI SON CV...

JE PEUX TRACER TOUT CE QU'IL A MANGÉ.
REMONTER SA CHAÎNE ALIMENTAIRE ET CELLE DES POULES, DES CŒURS ET DES LIMACES QU'IL A AVALÉS.

ET BIEN TÔT, ON POURRA TOUT Y METTRE ET TOUT CONTRÔLER.

SA CONSOMMATION DE CO₂.

SES POINTS RETRAITE
SON CASIER JUDICIAIRE

ÇA N'A PAS L'AIR MAL POUR LE JUGEMENT DERNIER...

AVEC ÇA, JE VAIS EN RÉCUPÉRER ENCORE PLUS EN ENFER.



Adoption de l'IA Act : un frein pour l'innovation européenne ?

Le Parlement européen a adopté le règlement encadrant l'intelligence artificielle et ses usages. Un texte qui implique des contraintes pour les entreprises du secteur, faisant craindre à certains acteurs des freins à l'innovation.

C'est le premier texte au monde à réguler l'intelligence artificielle. Négocié par les États membres en décembre 2023, l'IA Act a été approuvé le mercredi 13 mars par le Parlement européen. Le texte vise à assurer la sécurité contre les risques liés à l'intelligence artificielle (IA) et au respect des droits fondamentaux.

Concrètement, il interdit certaines applications liées à l'IA — par exemple, les systèmes de catégorisation biométrique utilisant des caractéristiques sensibles ; l'extraction non ciblée d'images faciales sur Internet ou par vidéo-surveillance dans le but de créer des bases de données de reconnaissance faciale ; la reconnaissance des émotions sur le lieu de travail et dans les établissements d'enseignement ; la police prédictive basée sur le profilage d'une personne et l'évaluation de ses caractéristiques ; la manipulation du comportement humain ou l'exploitation des vulnérabilités des personnes. En outre, les systèmes d'IA à usage général et leurs modèles devront respecter des règles de transparence, la législation européenne sur les droits d'auteur et la publication de résumés



détaillés des contenus utilisés pour leur entraînement. Quant aux systèmes plus puissants, ils devront se soumettre à des exigences supplémentaires — évaluation des modèles, évaluation et atténuation des risques systémiques, signalement des incidents... Les deep fakes devront être signalées comme telles. Entrant en vigueur vingt jours après sa publication, l'IA Act sera progressivement appliqué, sa pleine application prendra, en réalité, deux ans, avec des délais spécifiques pour certaines dispositions.

Un texte structurant pour le marché

Voilà pour les grandes lignes, mais qu'en est-il de l'application ? *« La loi sur l'IA est le point de départ d'un*

SENSIBILITÉ TECHNOLOGIQUE

Paul-Olivier Gibert estime que le principal risque est ailleurs. Dans le texte, « la définition de l'IA repose sur une liste de techniques de calculs et d'algorithmique en annexes. Ce n'est pas une définition aussi claire et nette que celle d'une donnée personnelle. Or, les réglementations qui ont cette sensibilité technique sont plus sensibles aux évolutions technologiques, contrairement au RGPD plutôt tourné vers l'application de principes généraux. » Une relative faiblesse « qui pourra être tranchée par la jurisprudence » et par des révisions régulières des obligations prévues par le texte. Un « bureau de l'IA » doit aussi aider les entreprises dans leur mise en conformité.

nouveau modèle de gouvernance fondé sur la technologie. Nous devons maintenant nous concentrer sur la mise en pratique de cette législation », a déclaré le co-rapporteur de la Commission des libertés civiles, Dragos Tudorache (Renew, Roumanie).

Principal enjeu ? Une application du texte qui n'entrave pas l'innovation. C'était l'inquiétude

du président de la République, Emmanuel Macron : que ledit texte freine le développement des startups européennes d'IA générative, notamment la société française Mistral AI. Des craintes que partagent de nombreux acteurs de la French Tech. Interrogé par *Les Echos*, Matthieu Rouif, cofondateur et CEO de l'application de retouche photo, Photoroom, déplorait « un mauvais texte pour l'innovation en Europe », car favorisant la création de valeur « pour les grandes entreprises tech et les startups en dehors du continent ». De son analyse, les LLM mettront aussi plus de temps à arriver sur le marché européen, soit une perte de compétitivité pour les entreprises européennes.

« Une crainte légitime », selon Paul-Olivier Gibert, président de l'AFCDP (Association Française des Correspondants à la protection des Données à caractère Personnel), qui nuance toutefois : « il est tout à fait possible que les choses n'évoluent pas en faveur des entreprises européennes. Mais lorsque l'on regarde par le passé, un texte comme le RGPD est devenu une référence mondiale pour la protection des données personnelles. » De son analyse, la loi sera structurante pour le marché et impactera l'offre des entreprises fournissant des services fondés sur l'IA, qui s'adapteront.

IA : Microsoft investit à coup de milliards en Espagne et en Allemagne

Après avoir mis sur la table une enveloppe record de 3,2 milliards de dollars (Mds\$) en Allemagne, la firme de Redmond se tourne vers l'Europe du Sud et a annoncé un investissement de 2,1 Mds\$ en Espagne au cours des deux prochaines années. Comme en Allemagne, il s'agit de son plus gros chèque depuis son arrivée dans le pays. Plus précisément, l'enveloppe correspond à une multiplication par quatre des investissements dans l'infrastructure IA et Cloud en Espagne. « *Notre investissement va au-delà de la simple construction de centres de données, il témoigne de notre engagement de 37 ans envers l'Espagne, sa sécurité, ainsi que le développement et la transformation numérique de son gouvernement, de ses entreprises et de ses citoyens* », a écrit Brad Smith sur X.

Une stratégie en quatre points

Microsoft ambitionne d'ouvrir une région cloud autour de Madrid ainsi qu'un campus de centres de données

en Aragon, dans l'est du pays, qui servira aux entreprises et entités publiques européennes. Car bien qu'américaine, Microsoft semble bien décidée à jouer la carte de la souveraineté si chère au vieux continent.

« *Cette infrastructure offrira aux services cloud de Microsoft la garantie maximale de sécurité, de confidentialité et de souveraineté des données, et mettra toute la gamme de solutions d'intelligence artificielle de l'entreprise à la disposition des entreprises et des administrations publiques espagnoles et européennes* », a détaillé la firme dans un communiqué.

Le président espagnol Pedro Sánchez a indiqué sur X avoir analysé avec Microsoft les opportunités de coopération pour renforcer la cybersécurité et promouvoir l'utilisation de l'intelligence artificielle dans l'administration publique. L'accord s'inscrit d'ailleurs dans le cadre de la Stratégie nationale de cybersécurité définie par le gouvernement espagnol qui s'articule autour de quatre points : l'extension de l'utilisation de l'IA dans l'administration



Brad Smith, Président de Microsoft.

publique afin de moderniser les processus administratifs, renforcer la cybersécurité nationale, améliorer la cyber-résilience des entreprises, et promouvoir une IA responsable.

Licenciements en cascade dans l'univers du JV

C'est une partie en mode hardcore qui se joue dans le secteur du jeu vidéo. L'éditeur américain Electronic Arts (EA) a annoncé licencier 5% de ses effectifs, ce qui correspond à environ 670 postes sur les 13 400 que compte l'entreprise. C'est la seconde restructuration en moins d'un an pour EA, qui avait déjà poussé vers la sortie 775 employés en mars dernier. Dans un courrier envoyé aux équipes de l'entreprise, le directeur général d'EA, Andrew Wilson, a justifié la décision par l'abandon de certains jeux et la volonté de « *nous éloigner du développement de futures licences dont nous ne pensons pas qu'elles réussiront dans notre secteur en évolution* ». Le développeur souhaite se recentrer sur ses licences phares et les jeux de sports.

Restructuration en pagaille

Cette nouvelle vague de licenciements intervient dans un contexte difficile dans l'univers du jeu vidéo, où restructuration et rationalisation semblent être les maîtres-mots. Mardi, c'est Jim Ryan, président & CEO de Sony Interactive Entertainment, aux manettes de la branche PlayStation, qui a annoncé se séparer d'environ 900 salariés, soit 8%

des effectifs. Sont concernés des salariés de studios mondialement connus tels qu'Insomniac Games (Spider-Man), Naughty Dog (The Last of Us) basés aux États-Unis, Guerilla (Horizon), ainsi que les équipes technologiques, créatives et d'assistance en Europe et au Royaume-Uni. « *Le studio londonien des PlayStation Studios fermera entièrement ses portes* », a également précisé Jim Ryan. Et comme chez EA, des projets ont été annulés. « *Nous avons examiné nos studios et notre portefeuille, évaluant des projets à différents stades de développement, et avons décidé que certains de ces projets n'iront pas de l'avant.* » En janvier, c'est Microsoft (Xbox) qui avait annoncé 1900 départs peu de temps après avoir (enfin) bouclé le rachat d'Activision Blizzard King (Call of Duty). La firme avait invoqué « *des domaines de chevauchement (avec Activision ndlr)* » pour justifier sa décision. Plus globalement, en additionnant tous les licenciements dans le secteur du jeu vidéo depuis janvier (Playtika, Riot Games, Electronic Arts...), ce sont 7 000 postes qui ont été supprimés, contre 9 000 sur toute l'année 2023, d'après une estimation du site VGC.

LPU via le cloud : Groq s'offre Definitive Intelligence

Deux ans après l'acquisition de Maxeler Technologies par l'entreprise californienne Groq, le spécialiste des LPU pour les calculs d'intelligence artificielle a de nouveau cassé la tirelire pour s'offrir cette fois, la startup Definitive Intelligence. Fondée en 2022 par Gavin Sherry et Sunny Madra, cette entreprise développe des solutions d'analyse de données et des chatbots.

Comme de nombreuses startups avant elle, Definitive Intelligence a su tirer profit de l'engouement généralisé pour la GenAI, et a bouclé un tour de table de 25 millions de dollars en 2022, suivi deux ans plus tard, par cette



opération de rachat, dont le montant n'a pas été divulgué. Groq, qui se positionne en tant que challenger de Nvidia sur le marché des GPU, conçoit des puces gravées en 14 nm capables d'exécuter des

LLM à plusieurs dizaines de milliards de paramètres. L'entreprise ambitionne de déployer sa plateforme GroqCloud, construite en collaboration avec l'équipe de Definitive Intelligence. Cette solution est présentée comme « un nouvel espace de développement pour les développeurs, avec une documentation entièrement intégrée, des exemples de code et un accès en libre-service. » La plateforme doit, en outre, faciliter

l'accès des entreprises clientes au moteur d'inférence Groq LPU™ via, pour le développement d'applications d'IA générative, sans nécessiter l'achat et l'implémentation de puces.

Zendesk reprend Ultimate

Fournisseur de solutions d'expérience client en ligne, Zendesk met la main sur Ultimate, une solution d'automatisation des services qui s'appuie sur l'intelligence artificielle.

Avec Ultimate, Zendesk souhaite proposer des agents dotés d'une intelligence artificielle améliorée dont l'approche n'est pas uniquement réactive, mais axée également autour de la résolution proactive des problèmes, en complément de l'expertise humaine.

Ces agents automatisent jusqu'à 80 % des demandes de support. Ils exploitent l'ensemble des sources de données à disposition et fournissent une personnalisation complète pour résoudre la majorité des cas d'utilisation, même les plus sophistiqués. Cette plateforme d'automatisation s'intègre à n'importe quel système backend. Les détails de l'acquisition ne sont pas connus.

No Code : Divalto met la main sur Flexio

Après le rachat des activités mobiles et collaboratives de RBS et l'acquisition de SwingMobility, Divalto continue sur sa lancée avec Flexio. Créé en 2016 et basé à Besançon, Flexio et son fondateur Julien Brugger constate à l'époque que les entreprises doivent choisir entre deux approches : d'un côté, les

solutions comme Excel, accessibles à tous, flexibles, mais aux fonctionnalités limitées. De l'autre, les solutions plus complètes, mais moins agiles et nécessitant des compétences en développement. De là est née l'idée du No Code Flexio, une solution de création de solutions métier sur-mesure sans qu'il ne soit nécessaire

d'avoir des notions en développement. Cette acquisition vise à simplifier encore plus l'utilisation des solutions de Divalto. Espaces de travail interactifs améliorés, traitement de données conversationnelles, gestion de droits avancée... compte parmi les nouvelles fonctionnalités à venir.

Zscaler croque Avalor

Zscaler s'offre des fonctionnalités d'analyse et la protection des données multisources. L'éditeur américain de solutions de cybersécurité vient de finaliser l'acquisition d'Avalor, une start-up israélienne spécialisée dans le domaine. Son produit phare, Data Fabric for Security est capable d'ingérer des données de plusieurs sources telles que les data lakes, les entrepôts de données, les bases de données SQL, les applications, etc. et de les analyser pour détecter les risques, les failles et les lacunes de conformité associées. Zscaler compte utiliser la technologie d'Avalor pour améliorer sa plateforme

Zero Trust Exchange qui permet aux entreprises de rendre leurs applications inaccessibles depuis le web public et de scanner le trafic réseau qu'elles reçoivent afin de rechercher des logiciels malveillants.

En avril dernier, la start-up qui possède des bureaux à New York et Tel Aviv avait réalisé un tour de table de 30 M\$ comprenant 25M\$ de série A et 5 M\$ d'amorçage. Si Zscaler n'a pas dévoilé le montant de l'acquisition, nos confrères de Globes avancent une somme de 350 millions de dollars.

Lutte contre l'ingénierie sociale : Anozr Way lève 6 millions d'euros

Récompensée du prix de la Startup au FIC 2023, Anozr Way trace sa voie. Après un premier tour de table de 2 millions d'euros (M€) en 2021, la jeune poussée rennaise a annoncé une nouvelle levée de fonds de 6 M€ en Série A. Cette opération a été menée par la banque d'affaires Clearwater International, avec le soutien du VC européen Hi Inov — Dentressangle et a été appuyée par Breizh Up et BNP Paribas Développement. Alors que le facteur humain est responsable de 80 % des cyberattaques, le principal risque pesant sur les entreprises est l'ingénierie sociale. C'est sur ce segment que se concentre la startup, qui a fait de la protection et de la maîtrise de l'empreinte numérique sa spécialité. Anozr Way reproduit la phase de reconnaissance des cybercriminels

pour identifier les personnes les plus exposées au sein d'une organisation. Les solutions qu'elle déploie permettent ensuite aux DSIs et RSSI d'évaluer et de superviser les risques, puis d'appliquer les mesures permettant de les réduire. Les collaborateurs peuvent également corriger eux-mêmes leur exposition cyber. Suite à cette montée en capital, Anozr Way prévoit de recruter 30 personnes et de pousser son développement à l'international. La priorité est donnée Québec, où elle ambitionne d'ouvrir des bureaux à Montréal, puis aux États-Unis.



Entourage lance un fonds early stage de 25 M€

Dédié à des entreprises créées sur le modèle SaaS, Entourage vise à construire et financer la prochaine génération d'entreprises SaaS B2B, en mettant l'accent sur les innovations en matière de produits et services qui vont stimuler la croissance et la transformation des entreprises. Le fonds est principalement composé d'entrepreneurs et le capital constitué sera utilisé pour créer et investir dans le

développement de nouvelles entreprises logicielles avec « Entourage Studio ». Une stratégie d'investissement Early stage ciblée viendra soutenir les entrepreneurs européens les plus ambitieux dans le domaine du SaaS. Le fonds a déjà commencé ses opérations avec Tekst.com, une jeune poussée qui développe une plateforme d'intelligence artificielle qui automatise l'assistance et

le back-office des grandes entreprises. Cette entreprise est en Incubation dans le fonds Studio. Entourage Capital se lance avec 3 sociétés SaaS B2B à forte croissance dans son portefeuille : Aikido, Henchman et NineID. Aikido développe des produits de cybersécurité. Henchman est une Legal Tech et NineID se spécialise sur les contrôles d'accès et gestion des sous-traitants.

Chiffrement homomorphe : Zama lève 73 millions de dollars

La startup de cryptographie spécialisée dans le développement de solutions de chiffrement homomorphe, Zama, a annoncé jeudi 7 mars 2024, avoir bouclé un tour de table de 73 millions de dollars en série A. L'opération a été dirigée par Multicoin Capital et Protocol Labs, avec la participation de Metaplanet, Blockchange Ventures, VSquared, Stake Capital, Juan Benet (fondateur de Filecoin), Anatoly Yakovenko (co-fondateur de Solana) et Gavin Wood (co-fondateur d'Ethereum et co-créateur de Polkadot). Avec ces fonds, la jeune poussée prévoit de renforcer son équipe de 75 personnes en recrutant de nouveaux ingénieurs,

développeurs logiciels et chercheurs en cryptographie. Ceci dans le but de maintenir ses bibliothèques open source pour les développeurs souhaitant concevoir des solutions FHE, et pour développer de nouveaux partenariats en vue de créer une nouvelle classe d'applications de chiffrement entièrement homomorphes (FHE). Cette technologie permet de traiter des données sans les déchiffrer. En d'autres termes, des entreprises peuvent fournir des services sans avoir besoin d'accéder aux données des utilisateurs et sans que cela n'affecte lesdits services.

Adaptive ML lève 20 millions de dollars



La startup franco-américaine Adaptive ML a bouclé un tour de tables en seed à 20 millions de dollars dirigé par Index Ventures ICONIQ Capital, Motier Ventures, Databricks Ventures, IRIS, HuggingFund by Factorial. La jeune poussée veut proposer des modèles d'IA adaptés aux cas d'usage des entreprises et en apprentissage perpétuel grâce aux interactions des utilisateurs.

« Chez Adaptive ML, nous rationalisons la voie vers des applications genAI plus intuitives et plus performantes, améliorant directement l'expérience utilisateur et les résultats commerciaux », a déclaré le PDG Julien Launay. La startup prévoit d'ouvrir un centre de R&D à Paris et de renforcer ses effectifs.

Sopra Steria renforce sa collaboration avec Microsoft

L'ESN intègre le programme Early Access de Copilot for cybersecurity de Microsoft et vise à devenir un acteur de référence dans le secteur.

Sopra Steria a choisi de faire partie dès maintenant des premières entreprises à bénéficier du programme Early Access de Microsoft Copilot for Security, la solution de cybersécurité de Microsoft basée sur l'IA et permettant aux professionnels de la sécurité de réagir rapidement aux cyberattaques, de traiter les signaux avec la rapidité de la machine et d'évaluer l'exposition aux risques en quelques minutes. Cet accès privilégié aux dernières capacités de l'IA générative appliquée à la cybersécurité, ainsi qu'un accompagnement personnalisé de la part des équipes de Microsoft, a été intégré dans le cadre du programme IA/cyber de Sopra Steria.

Microsoft Copilot for Security combine la puissance de GPT-4 avec un modèle spécifique à la sécurité de Microsoft couplé à des capacités d'apprentissage du comportement des analystes cyber, ce qui permettra de générer des données et des recommandations personnalisées pour les équipes de Sopra Steria, en réponse à des requêtes formulées en langage naturel sur l'ensemble de son offre cybersécurité : prévention, protection, détection & réponse aux cyber-attaques. Les équipes cybersécurité de Sopra Steria travaillent en synergie avec les équipes de développement de Microsoft Copilot for Security dans le cadre de ce programme.

Databricks dans le sens du vent avec Mistral AI

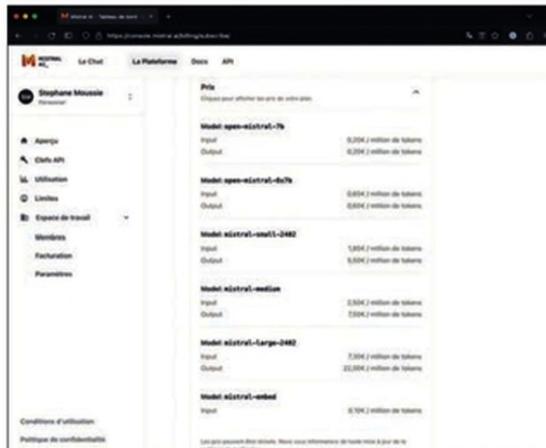
Databricks annonce un partenariat et une participation au financement de série A de Mistral AI.

Databricks et Mistral proposent désormais les modèles ouverts de Mistral intégrés nativement au sein de la « Data Intelligence Platform » de Databricks. Les clients de Databricks peuvent désormais accéder aux modèles de Mistral AI sur la Marketplace de Databricks, interagir avec ces modèles dans le Mosaic AI Playground, les utiliser comme endpoints de modèles optimisés grâce à Mosaic AI Model Serving, et les personnaliser sur leurs propres données grâce à l'adaptation.

Les clients de Databricks pourront utiliser et adapter les modèles de Mistral en utilisant leurs propres données pour développer une application. Grâce à ces intégrations clés en main, les entreprises peuvent rapidement exploiter les modèles de Mistral pour leurs applications d'IA générative tout en bénéficiant des bénéfices de la plateforme de Databricks en termes de sécurité, de confidentialité et de gouvernance des données.

Deux principaux modèles sont mis en avant le 7B et le 8x7B. Le premier est un modèle de transformer dense, petit mais puissant, conçu avec une longueur de fenêtre contextuelle de 8k. Il est très efficace en raison de sa taille relativement petite de 7 milliards de paramètres et de son architecture de modèle qui tire parti de l'Attention aux requêtes groupées (GQA) et de l'Attention à fenêtre coulissante (SWA).

Le second est un modèle de SMoE (Sparse Mixture of Experts), prenant en



Les prix des différents modèles de Mistral AI.

charge une longueur de fenêtre contextuelle de 32k et capable de traiter l'anglais, le français, l'italien, l'allemand et l'espagnol. Il surpasse Llama 2 70B sur la plupart des benchmarks, tout en offrant une inférence 6x plus rapide grâce à son architecture SMoE qui n'active que 12 milliards de paramètres pendant l'inférence, sur un total de 45 milliards de paramètres entraînés.

La personnalisation se réalise par interaction avec le Mosaic AI Playground en utilisant les modèles comme des endpoints de modèles optimisés grâce à Mosaic AI Model Serving. Mosaic AI Foundation Model APIs est une fonctionnalité de diffusion de modèles (Model Serving) qui permet aux clients d'accéder à Mistral 8x7B (ainsi qu'à d'autres modèles de pointe) et de lancer des requêtes, en tirant parti de déploiements de modèles

hautement optimisés et sans avoir à créer et à maintenir des déploiements et des endpoints. Avec Databricks Mosaic AI Model Serving, les clients peuvent accéder aux modèles de Mistral AI à l'aide des mêmes API que celles utilisées pour les autres modèles de fondation (Foundation Models). Cela permet aux clients de déployer, gouverner, interroger et surveiller n'importe quel modèle de fondation à travers les clouds et les prestataires,

ce qui rend possibles l'expérimentation et la mise en production de grands modèles de langage. Les clients peuvent également faire appel à l'inférence de modèle directement à partir de Databricks SQL en utilisant la fonction SQL `ai_query`.

Pour expérimenter rapidement les modèles Mistral pré-entraînés et affinés, les clients peuvent accéder au Mosaic AI Playground disponible dans la fenêtre Databricks. L'AI Playground permet des conversations interactives multi-tours, l'expérimentation des paramètres d'échantillonnage de l'inférence de modèle tels que la température et max_tokens, et l'inférence côte à côte de différents modèles pour observer la qualité de la réponse du modèle et les caractéristiques de performance.

ABYY et Novelis resserrent leurs liens

Après des premiers succès commerciaux, les deux entités renforcent leur coopération pour déployer des projets d'IA en Europe et aux États-Unis.

Novelis a optimisé les plateformes de litiges électroniques du Conseil d'État ainsi que celles d'Abeille Assurances en s'appuyant sur SS&C Blue Prism et sur ABYY Process Intelligence. Cette initiative a permis l'automatisation efficace des tâches administratives répétitives, tout en identifiant les priorités de développement essentielles dans ces entreprises. Cette démarche soutient activement la transformation digitale des systèmes d'information existants de ces organisations,

contribuant ainsi à renforcer leur agilité dans un environnement où la résilience est de plus en plus importante.

Par ailleurs, Novelis a réussi à déployer avec succès les solutions de traitement intelligent de documents (IDP) d'ABYY pour optimiser la saisie de données au sein d'une société de gestion de créances de premier plan en France. Cette initiative a conduit à une réduction significative des erreurs potentielles, évitant ainsi les retards dans les processus

entraînant l'invalidation des réclamations. Les principaux bénéfices de cette mise en œuvre sont l'éradication des erreurs humaines, une réduction de 97 % du temps de traitement qui est passé de 45 minutes à seulement 1,5 minute en moyenne. Les gains de productivité sont l'équivalent de 5 équivalent temps plein et a permis une allocation renforcée des ressources aux processus en lien avec l'expérience client.

Datafolio et Dott signent un partenariat

Les deux entités ont signé un partenariat de gestion des risques de la mobilité et d'intermédiation d'assurances couvrant depuis le 1^{er} février 2024 les véhicules et utilisateurs de la plateforme Dott en France, en Italie et en Espagne.



Une vue de la proposition d'une assurance de datafolio sur mobile.

La jeune insurtech franco-italienne basée à Lille et cofondée par Yuri Narozniak, développe depuis 2020 des solutions technologiques et des contrats d'assurance répondant aux besoins liés au fort développement des nouveaux modèles et comportements de mobilité.

Après un lancement réussi fin 2022 en Italie avec un premier partenaire, Bit Mobility, la société a rapidement su convaincre RideMovi (Italie, Espagne, Portugal et Suède) et désormais Dott, l'un des leaders européens présent dans 7 pays et opérant plus de 50 000 véhicules. La solution Datafolio permet non seulement l'automatisation des opérations liées à la gestion des contrats d'assurance, mais aussi, grâce à des modèles combinant science actuarielle, science des données et intelligence artificielle, aide les plateformes à minimiser leur exposition aux risques par l'analyse de leurs données et la prévention. De plus, la suite technologique EI. OS (pour « *Embedded Insurance Operating System* ») de Datafolio permet aux plateformes de mobilité de distribuer des assurances complémentaires à leurs utilisateurs de façon totalement intégrée dans leur parcours client et 100 % à l'usage, générant ainsi des revenus supplémentaires significatifs.

AGENDA

Adobe Summit
26-28 mars 2024

Venetian Convention and Expo Center, Las Vegas, Nevada, USA

Systèmes et objets Connectés + Cloud Forum
3-4 avril 2024
Paris, Porte de Versailles

Google Cloud Next
9-11 avril 2024
Mandalay Bay, Las Vegas, Nevada, USA

MPLS SD & AI Net World Congress
9-11 avril 2024
Palais des Congrès, Pte Maillot, Paris

Paris Blockchain Week
8-12 avril 2024
Carrousel du Louvre, Paris

NAB Show
13-17 avril 2024
Las Vegas Convention Center, Las Vegas, USA

RSA conference
6-9 mai 2024
Moscone Center San Francisco

Microsoft Build Conférence hybride
21-23 mai 2024
Seattle et en ligne

Protéger votre entreprise de l'edge au cloud. **Un acte responsable.**

Protégez vos données avec le bon niveau de sécurité.

Combler les lacunes et diminuer la complexité, de l'edge au cloud et partout entre les deux.

Découvrir greenlake.hpe.com 





Blockchain

Le renouveau

Après une vague d'engouement à la fin de la dernière décennie, la technologie de blockchain a connu une période creuse pendant plusieurs années et le grand public s'est concentré sur certaines chaînes publiques et les cryptomonnaies. Aujourd'hui après cette période de déception les cas d'usages et de nouvelles possibilités qui abaissent le coût de gestion d'une chaîne remettent la technologie au goût du jour et les projets se multiplient. D'ailleurs les cabinets d'analystes parient encore sur la chaîne de blocs et lui prévoit d'ici peu un avenir brillant avec de nombreuses applications en entreprises.

Sous de meilleurs auspices

Depuis quelques mois la chaîne de blocs connaît une sorte de renouveau après avoir connu une période creuse avec des utilisateurs proches de la désillusion malgré les promesses de la technologie. Les analystes et les entreprises semblent avoir retrouvé de l'appétence et des projets ressurgissent.

Pour ceux qui ne sont pas familiers avec cette nouvelle technologie, la Blockchain ou chaîne de blocs se définit comme un registre partagé et non modifiable qui facilite le processus d'enregistrement des transactions et de suivi des actifs dans un réseau d'entreprises. Son but est de fournir des informations immédiates, partagées et totalement transparentes, stockées dans un registre non modifiable auquel seuls les membres autorisés du réseau peuvent accéder.

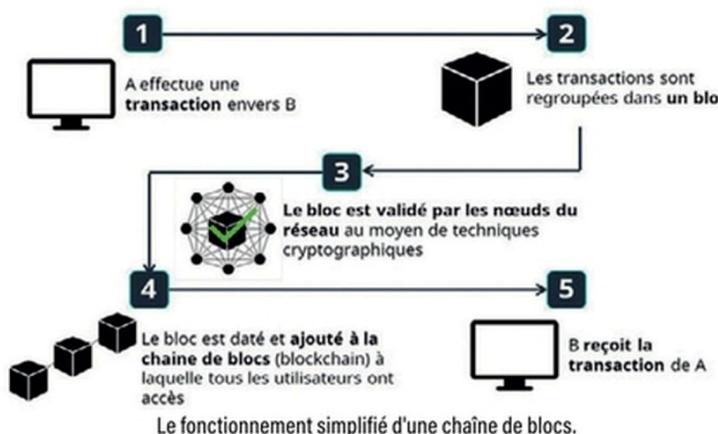
Plusieurs types de blockchain

Il existe différentes variantes de cette technologie. La première est la plus connue : la blockchain publique. Quiconque peut la rejoindre pour y participer. La puissance de calcul nécessaire, le caractère peu ou pas confidentiel des transactions et le faible niveau de sécurité peuvent être des inconvénients. Ces aspects sont à prendre en compte dans les cas d'utilisation de la blockchain par les entreprises. À l'instar d'un réseau de blockchain public, un réseau de blockchain privé est un réseau décentralisé de pair à pair. Cependant, une organisation contrôle le réseau, en déterminant qui est autorisé à participer, à exécuter un protocole de consensus et à gérer le registre partagé. Selon le cas d'utilisation, cela peut renforcer significativement la confiance entre les participants. Une blockchain privée peut être exécutée derrière un pare-feu d'entreprise et même hébergée sur site. Les entreprises qui mettent



Le manifeste de Michelin sur son club.

en place une blockchain privée créent généralement un réseau de blockchain autorisé. Il est important de noter que les réseaux de blockchain publics peuvent également être soumis à des autorisations. Dans ce cas, des restrictions sont appliquées pour définir qui peut participer au réseau et spécifier les transactions dans lesquelles le participant peut être impliqué. Les participants doivent obtenir une invitation ou une autorisation pour joindre le réseau. Plusieurs organisations peuvent se partager les responsabilités de la gestion d'une blockchain. Ces organisations présélectionnées déterminent qui peut soumettre des transactions ou accéder aux données. Une blockchain de consortium est idéale pour les entreprises lorsque tous les participants doivent être autorisés et avoir une responsabilité partagée pour la blockchain. La tendance actuelle est de mixer différents types de blockchain qui tend vers des modèles hybrides entre chaîne publique et privée. Les chaînes privées sont cependant peu à peu délaissées du fait de leur



coût de fonctionnement et de la puissance comme calcul nécessaire pour réaliser les opérations. De plus, elle ne correspond pas à de nombreux cas d'usages ce qui limite son utilisation possible.

Un marché qui va exploser

Selon Fortune Business Insights, le marché global de la blockchain était de 11,14 milliards de dollars en 2022 et plus de 17 milliards de dollars l'année dernière. En 2030, l'institut prévoit un marché de 469,49 milliards de dollars avec un taux de croissance pondérée par an de 59,9% durant la période considérée. La recrudescence des attaques contre les données et la digitalisation d'industries de plus en plus diverses vont être les catalyseurs de ce développement. Cette dernière tendance pousse de plus l'utilisation de services de BaaS (Blockchain as a Service). Le secteur de la finance et en particulier le secteur des paiements va conduire durant les prochains mois ainsi que la zone Amérique du Nord si l'on considère la croissance sous un angle géographique sous l'impulsion de fonds d'investissements pour apporter de l'innovation dans la finance décentralisée, la cybersécurité et le commerce en ligne.

Des usages ciblés en entreprise

Malgré une certaine désillusion sur la technologie, les entreprises continuent à s'intéresser à la chaîne de blocs et explorent les usages qui leur permettraient d'être plus efficaces et plus sécurisées comme la gestion des identités et des accès, la gestion de la chaîne d'approvisionnement, les contrats intelligents (smart contracts), la gestion documentaire et la vérification d'opération. Autre usage à signaler, la conformité en particulier la conformité RSE pour s'assurer de la provenance ou de la véritable qualité de produits dans leur chaîne d'approvisionnement.

L'émergence des tokens

Si un bon nombre d'entreprises s'interrogent encore sur les cas d'usages qu'elles pourraient faire de la chaîne de blocs, elles sont déjà un certain nombre à se convertir aux NFT (Non Fungible Tokens) pour créer de nouvelles sources revenus en vendant des produits ou des actifs digitaux par ces jetons. Ainsi Michelin s'est lancé dans les NFT avec le Michelin 3xplorer Club qui propose 5000 NFT à l'effigie de Bibendum sur la plate-forme Ethereum. Les détenteurs des NFT pourront avoir des expériences exclusives et vivre de l'intérieur les activités de Michelin.

Le frein des coûts

Malgré cette appétence pour la technologie, peu d'entreprises en dehors du secteur de la finance voient des gains assez significatifs pour remplacer la solution qu'elles détiennent actuellement. En ce sens l'offre du marché doit s'améliorer pour démontrer aux entreprises que la blockchain est supérieure à ce que les entreprises utilisent actuellement par la création de logiciels et de services qui aident véritablement les entreprises. □

B.G

FLEUR PELLERIN REJOINT LE CONSEIL DE TEZOS FOUNDATION



La Tezos Foundation basée en Suisse sous l'autorité d'une institution de supervision fédérale promeut l'utilisation de la plate-forme de chaîne de blocs Tezos, une plate-forme open source. Si la fondation ne gère pas en direct la plate-forme, elle est ouverte sur l'écosystème autour de la plate-forme mais aide à son développement et à sa reconnaissance en appuyant les entreprises utilisant Tezos. La fondation gère une soixantaine de projets et a approuvé plus de 18 millions de dollars de fonds sur ceux-ci.

Le conseil que rejoint Fleur Pellerin a pour tâche de gérer et d'administrer les actifs de la fondation et d'en définir la stratégie. Le conseil délègue de plus la gestion opérationnelle et des actifs au comité exécutif de la fondation. Le conseil choisit ses membres et est composé de huit personnes. Depuis 2016, Fleur Pellerin dirige Korelya Capital, le fonds de capital-risque qu'elle a créé, dédié aux entreprises de l'écosystème tech européen. Déjà impliquée dans le secteur du Web3, via ce fonds, Fleur Pellerin a notamment contribué au développement d'entreprises comme Ledger, Bitfury et Ramp. Son expérience acquise lors de ses précédentes fonctions dans la culture et le numérique, ainsi que son expertise dans l'investissement, sont vues comme de véritables atouts pour le projet.

Vers les blockchain hybrides

Après des débuts en boulet de canon la blockchain a connu une période plus creuse. Aujourd'hui les entreprises ont augmenté leur maturité et se tournent vers des technologies différentes des temps héroïques avec des choix vers les plates-formes hybrides ou qui simplifient largement l'utilisation des chaînes de blocs.

Pour Vincent Fournier, Blockchain & Digital Assets Leader chez IBM, précise : « la difficulté provient que le mot Blockchain englobe beaucoup de choses, de solutions de traçabilité alternative avec plus de décentralisation et horodatage. De plus les environnements crypto ont beaucoup évolué, c'est donc un mot difficile à comprendre avec des volets techniques différents ».

Un cheminement technologique

Si les chaînes de blocs publiques sont désormais assez anciennes, elles ne satisfaisaient pas totalement les entreprises à leurs débuts. Come le rappelle Vincent Fournier : « comme à chaque fois il y a une phase d'effervescence, on teste, on publie, on réalise des études de cadrage, des PoC (Proof of Concept) et parfois sur des mises en production ». Les premiers essais ont donc principalement concernés la chaîne logistique allant au-delà de la simple traçabilité en ajoutant le lignage des produits depuis leur source tout en prenant en compte des éléments de volumétrie. Des entreprises comme Carrefour se sont lancées rapidement sur ce domaine et quasiment tout la nomenclature de produits du détaillant sont aujourd'hui sont tracées avec la chaîne de blocs. Si ce projet a abouti, plusieurs autres n'ont pas connu le même destin avec des abandon ou

Vincent Fournier,
Blockchain &
Digital Assets Leader
chez IBM.



« La difficulté provient que le mot Blockchain englobe beaucoup de choses. »

Emmanuel Méthivier,
Business Program
Director chez Axway.



« Les grands comptes sont un peu tombés dans le panneau de vouloir décentraliser pour recentraliser sur ces chaînes privées. »

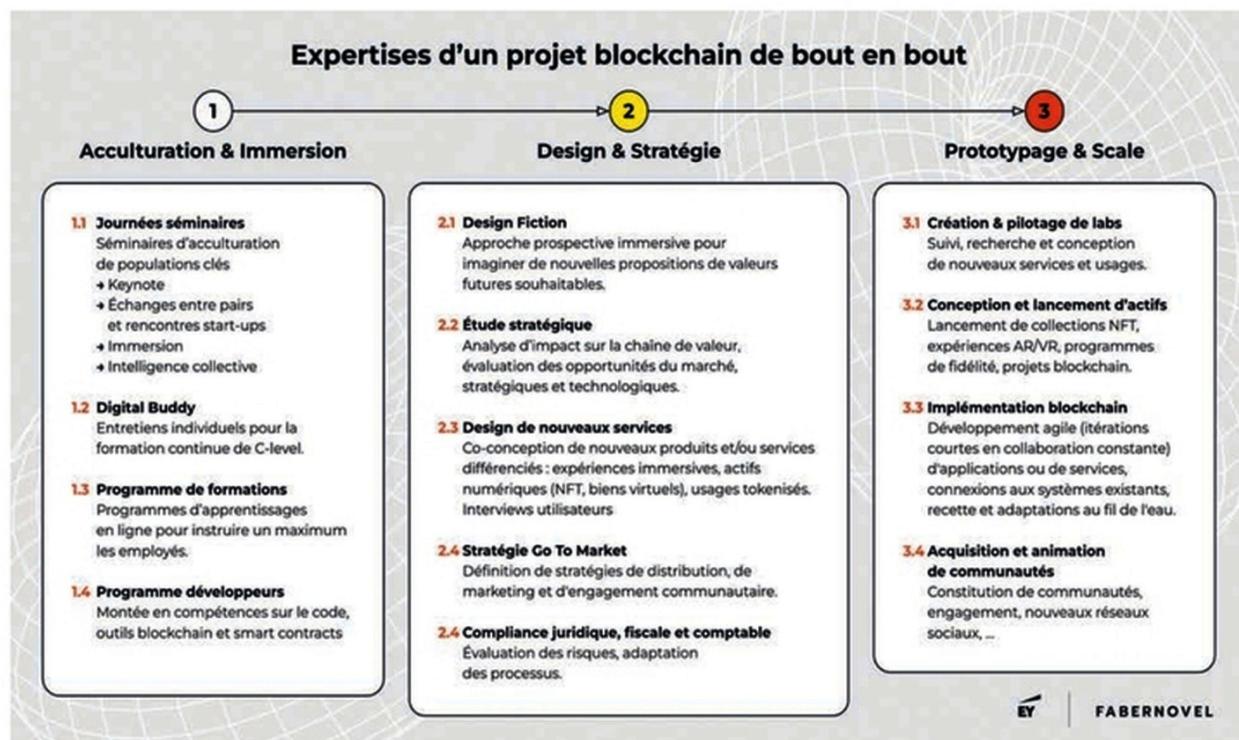
des retours en arrière selon la maturité des clients. Vincent Fournier le concède : « tous les arguments n'étaient pas là en termes d'anonymat, de confidentialité et de performances ». Il ajoute : « cette phase d'exploration n'a pas levé le doute ».

Emmanuel Méthivier, Business Program Director chez Axway, va plus loin : « de grands éditeurs se sont emparés du sujet en proposant une superbe technologie motorisée par des crypto-monnaies avec des concepts de Proof of Work (Preuve de travail). C'est ainsi que fonctionne le Bitcoin, plus vous utiliser la plate-forme et plus vous gagner de jetons (tokens) qui sont valorisés. Il est apparu la blockchain privée qui avait pour promesse de profiter des bénéfices de la blockchain sans mettre les mains dans le cambouis. Les grands comptes sont un peu tombés dans le panneau de vouloir décentraliser pour recentraliser sur ces chaînes privées ».

Arthur de la Brunière, Manager Blockchain chez EY Fabernovel, renchérit : « les coûts de production de l'infrastructure nécessaire étaient trop importants. Mersk dépensait ainsi 8,5 M/ mois juste pour cela. Ils ont abandonné le projet ». Il ajoute : « les entreprises se tournent maintenant vers des chaînes plus ouvertes (publiques) ou hybrides ».

Les chouchous du moment

Les blockchains hybrides combinent des éléments de chaînes publiques et des éléments de réseaux privés.



Les principales étapes et expertises à posséder sur un projet blockchain.

Polygon (anciennement Matic Network) est une solution de Layer 2 / sidechain qui permet aux utilisateurs d'accéder aux applications Ethereum (DApps) avec des frais réduits et une latence quasi nulle.

Dans sa recherche, l'équipe de Matic Network tombe sur une publication du co-fondateur d'Ethereum, Vitalik Buterin et de Joseph Poon, co-fondateur du Lightning Network, détaillant une nouvelle solution de scalabilité : « Plasma ». Cette solution repose sur la duplication d'une blockchain mère, dont la copie sera plus rapide tout en conservant la sécurité. La solution Plasma formera la base de la blockchain Matic. Matic Network devient Polygon au début de l'année 2021. Bien plus qu'un changement de

nom, cette refonte signe une réorientation d'une unique blockchain vers un agrégateur de solutions de scalabilité pour Ethereum. N'importe quel développeur peut donc venir y créer une blockchain scalable. Polygon fonctionne en connectant les DApps Ethereum à un réseau de sidechains reliées entre elles. Les utilisateurs peuvent donc interagir avec les DApps sur le réseau Polygon sans avoir à attendre les confirmations de bloc sur le réseau Ethereum principal. La particularité principale de Polygon est sa capacité à gérer plusieurs millions de transactions par seconde. Cela est possible grâce à l'utilisation de Plasma et d'autres technologies de pointe telles que l'optimisation du protocole de preuve-à-enjeu (ou Proof-of-Stake) et des Shard Chains. En résumé, Polygon permet aux utilisateurs d'accéder aux applications Ethereum avec des frais réduits et une latence quasi nulle. De plus, le réseau est capable de gérer plusieurs millions de transactions par seconde.

Chaque sidechain est une blockchain indépendante qui traite les transactions et les smart contracts de manière autonome. Cependant, ces sidechains sont interconnectées avec la blockchain Ethereum, ce qui permet la transférabilité des actifs entre les différentes chaînes.

L'un des aspects fondamentaux de Polygon est sa capacité à permettre aux utilisateurs de transférer leurs actifs entre la blockchain principale d'Ethereum et les sidechains de Polygon. Cela se fait via des ponts (bridges) qui sécurisent les actifs lorsqu'ils traversent entre les deux réseaux. Polygon propose différents types de sidechains adaptées à divers cas d'utilisation. Parmi elles, on trouve les sidechains PoS (Proof of Stake) qui utilisent un consensus en preuve d'enjeu pour valider les transactions.

Arthur de la Brunière, Manager Blockchain chez EY Fabernovel.



« Les entreprises se tournent maintenant vers des chaînes plus ouvertes (publiques) ou hybrides. »

Il y a également les sidechains PoA (Proof of Authority) où un groupe restreint d'entités est responsable de la validation des blocs. Chaque type de sidechain a ses avantages et est adapté à des cas d'utilisation spécifiques.

Polkadot est un autre protocole qui soulève l'intérêt. La solution est un réseau composé de plusieurs blockchains interconnectées et sert de fondement à la création d'un ensemble de blockchains interopérables,

où l'efficacité de leurs communications est mêlée à une forte sécurité et transparence. Le principal intérêt de Polkadot est de permettre des communications fluides et sécurisées entre plusieurs blockchains dans un seul et même écosystème sans avoir besoin d'utiliser de « ponts blockchain » (bridges). Cela évite les problèmes techniques liés au Bridge qui sont souvent sujets à des attaques. □

B.G

Le terrain de jeu des start-up

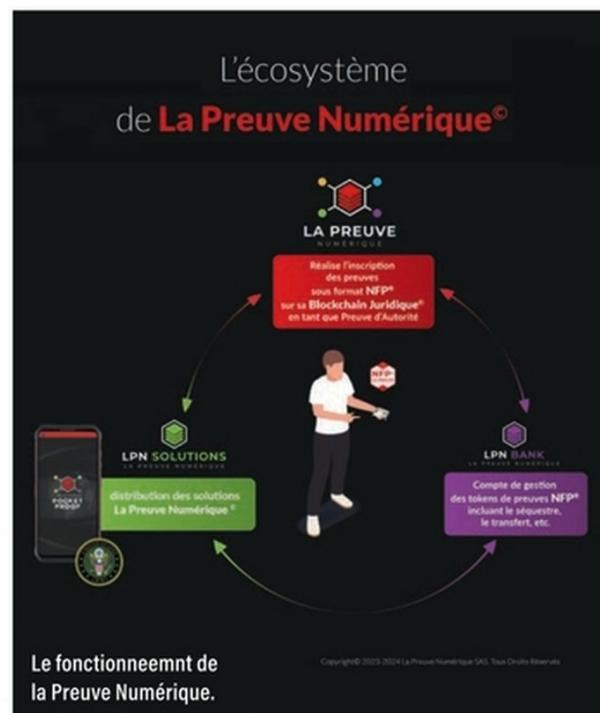
La technologie de Blockchain a de multiples cas d'usages possibles. De nombreuses start-up se sont donc engouffrées sur elle avec souvent beaucoup d'originalité... et avec peut-être l'arrière-pensée que le secteur est très convoité par les investisseurs.

Ta Da, à l'origine Vivoka, développe des modèles d'assistants vocaux développés sur de l'apprentissage machine pour la reconnaissance vocale. Les bots sont en plusieurs langues. La société connaît des difficultés pour trouver des jeux de données pour entraîner les modèles dans des langues rares. De plus « la qualité n'était pas toujours au rendez-vous » explique Valentin Thauron, PDG de l'entreprise. « Nous avons donc décidé de créer notre propre plate-forme de collecte » ajoute-t-il. Ainsi La plate-forme mais en relation des entreprises avec des utilisateurs de partout dans le monde enregistrés sur la plate-forme. Par exemple, une entreprise souhaite un robot de reconnaissance vocale en roumain, ils vont sur la plate-forme et distribue le travail auprès des utilisateurs répondant à leurs critères enregistrés sur la plate-forme. La candidature est vérifiée et retournée au client. Ta Da s'assure que ce sont bien des locuteurs natifs. De plus celui-ci peut ou non accepter la rémunération proposée pour le travail ce qui équilibre le marché par les prix.

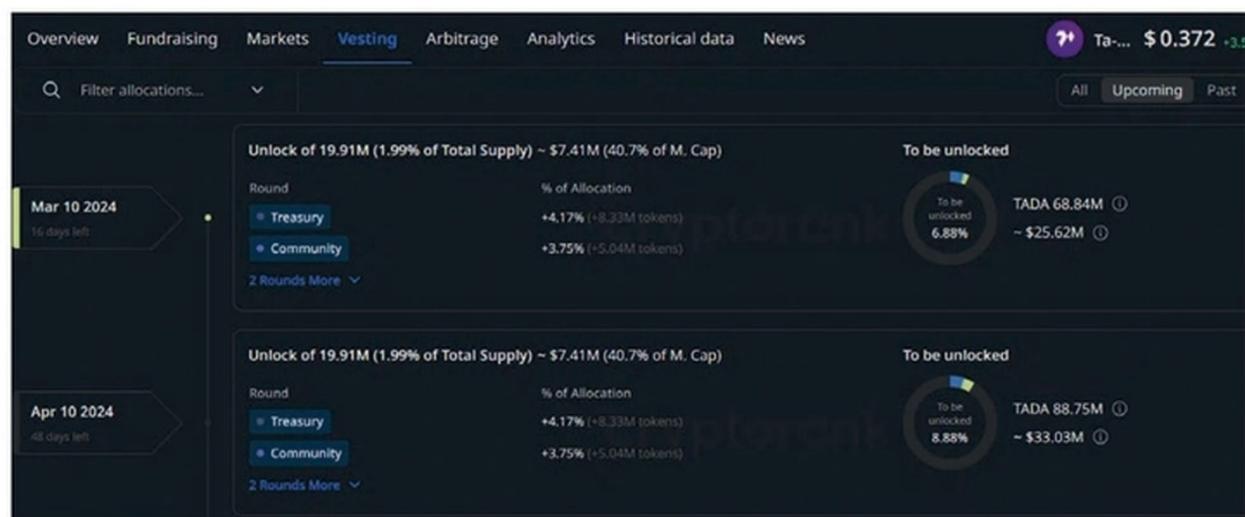
La blockchain est là pour assurer la traçabilité sur les vérifications. « Auparavant nous n'avions aucune visibilité et 30 à 40 % des utilisateurs ne correspondaient pas au souhait des entreprises » explique le CEO de Ta Da. « Grâce à la blockchain on peut tout tracer. Les entreprises ont les preuves du consensus et de la fiabilité de la donnée ». L'entreprise ne veut pas s'arrêter sur les liens vers 6 chaînes publiques actuellement en place et vise à développer sa propre instance sur Cosmos SDK, un substrat de Polkadot.

La preuve numérique

Autre exemple, la Preuve Numérique a été créé à partir d'un cas concret d'avocat. Christophe Curtelin, président et co-fondateur de la Preuve Numérique, raconte : « une



entreprise française avait un contentieux avec une entreprise chinoise pour un contrat non exécuté de construction de machines. Tous les éléments de preuves et la reconnaissance des preuves numériques ont permis de gagner le procès. Sur notre chaîne propriétaire par conception, Didier Urban a intégré les critères de preuves juridiques numériques reconnues internationalement. Nous sommes un spécialiste de la preuve, une ProofTech » assume fièrement Christophe Curtelin. Il continue : « notre but est de construire un dossier de preuve



La création de jeton dans Ta Da.

le plus précis possible. Nous avons ainsi recodé les brevets et sur les principes de preuve nous apportons des jetons, des NFT échangeables entre partenaires». La deuxième difficulté était de faire que les pièces soient immédiatement recevables quelque soit la juridiction. Les jetons servent à cet effet pour la reconnaissance directe du jeton de preuve américain ou d'une autre nationalité.

De plus pour assurer la chaîne de preuve, l'entreprise a développé un appareil mobile « pocket proof » qui permet de saisir la donnée au plus près qui se connecte à un boîtier, « station proof » pour ne laisser aucun doute sur la preuve. L'information est captée sur le terminal qui embarque la chaîne de blocs de la Preuve Numérique et devient impossible à modifier.

Un mode de paiement alternatif

Fipto est une solution de paiements B2B à l'international permettant aux entreprises d'envoyer et de recevoir instantanément des devises fiat et numériques partout dans le monde via les rails de la blockchain. L'entreprise combine les meilleurs standards des paiements traditionnels avec la rapidité et la transparence de la technologie blockchain, offrant ainsi à ses utilisateurs une plateforme respectant les exigences de conformité pour des paiements cross-border rapides, sécurisés et disponibles 24h/24 et 7j/7. L'entreprise a obtenu son agrément PSAN en mars 2022. La solution ne propose pas son propre protocole mais interagit avec les protocoles existants comme Solana si le besoin de performance est important, ou Ethereum ou d'autres protocoles si le client le demande.

Le paiement en monnaie digital devient un enjeu stratégique pour toutes les entreprises ; c'est pourquoi Fipto leur permet d'exécuter facilement et en toute conformité, TOUS leurs paiements (en monnaie digitale et FIAT) en combinant les meilleures pratiques du paiement traditionnel et le meilleur de la technologie blockchain. Fipto permet aux entreprises de payer dans TOUTES les devises (environ 10 monnaies digitales) et de gérer tous les types de transactions (Fiat / crypto, Crypto / crypto, Crypto / fiat)

La plate-forme permet à toutes les entreprises d'interagir avec des services financiers portés par des technologies Blockchain d'une manière conforme à celles de leur département finance.

L'entreprise commence seulement la commercialisation de sa plate-forme. L'idée est de se diriger vers un système de facturation mensuel qui prenne en compte la volumétrie de transaction des clients et leur utilisation des APIs (cf modèle de facturation BaaS) ainsi que des frais transactionnels (change / paiement).

Des identités décentralisées

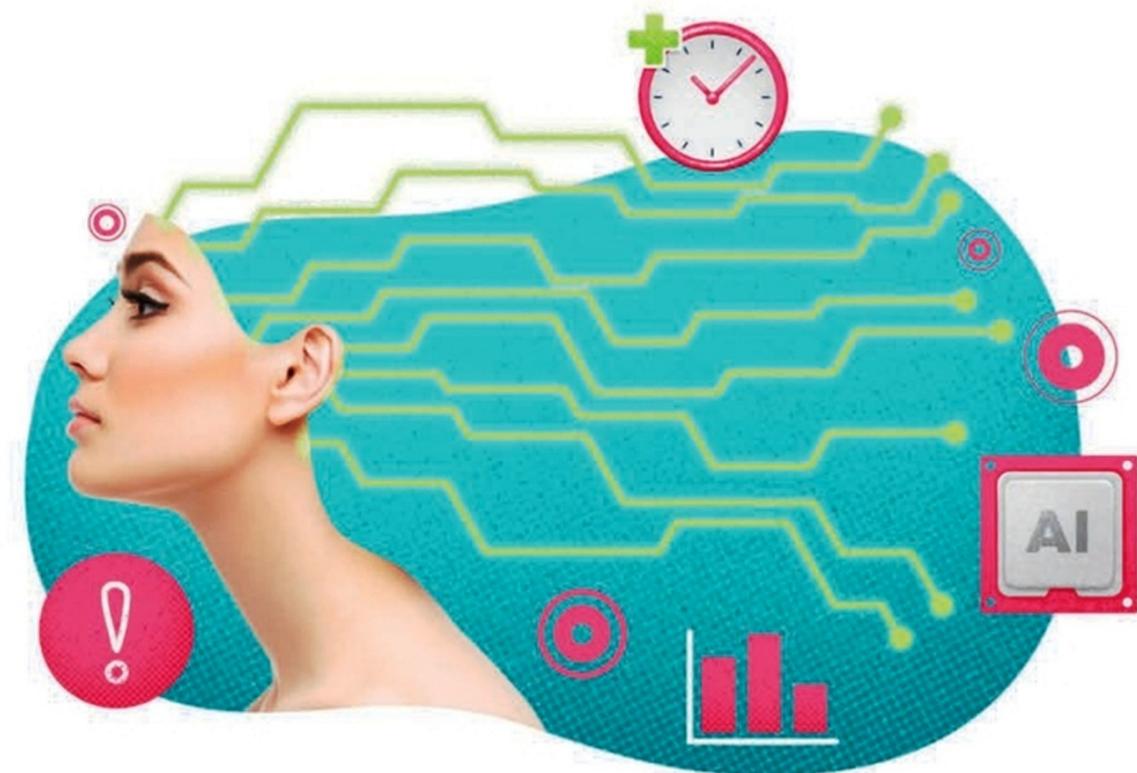
Archipels est issue d'un consortium de La Poste, Engie, EDF, la Caisse des Dépôts et INGroupe et dispose d'une blockchain privée et permissionnée. La société proposons un ensemble de services de confiance fonctionnant grâce à la blockchain notamment pour ses principes de transparence, confidentialité et sécurité. La solution comporte un wallet d'identité décentralisé (utilisant donc la blockchain) qui répond à la réglementation votée ce mois-ci par la Commission Européenne (eIDAS2). L'objectif de ce portefeuille numérique sécurisé est de libérer le potentiel de l'identité numérique ; notamment de fluidifier et sécuriser tous les échanges entre entreprises et individus. L'idée est que demain (2027), chaque citoyen européen aura son portefeuille numérique ultra-sécurisé qui contiendra son identité et toutes ses attestations/données numériques qu'il souhaite. Ainsi, les citoyens pourront reprendre un contrôle total de leurs données et choisir quand les partager et à qui (preuve de majorité, IBAN, diplôme...).

Les entreprises, elles, peuvent demander la présentation des attestations numériques dont elles ont besoin et fluidifier leurs process tout en se protégeant de la fraude et l'usurpation

Clairement, la technologie de la chaîne de blocs peut servir de nombreux secteurs et usages. De quoi renforcer son renouveau dans les mois à venir. □

B.G

PRÊT À TRANSFORMER LA GESTION DE VOS SERVICES INFORMATIQUES ?



50%

de réduction
des coûts

25%

d'amélioration du
temps de résolution
des tickets

20%

d'incidents en moins
grâce au monitoring
prédictif

70%

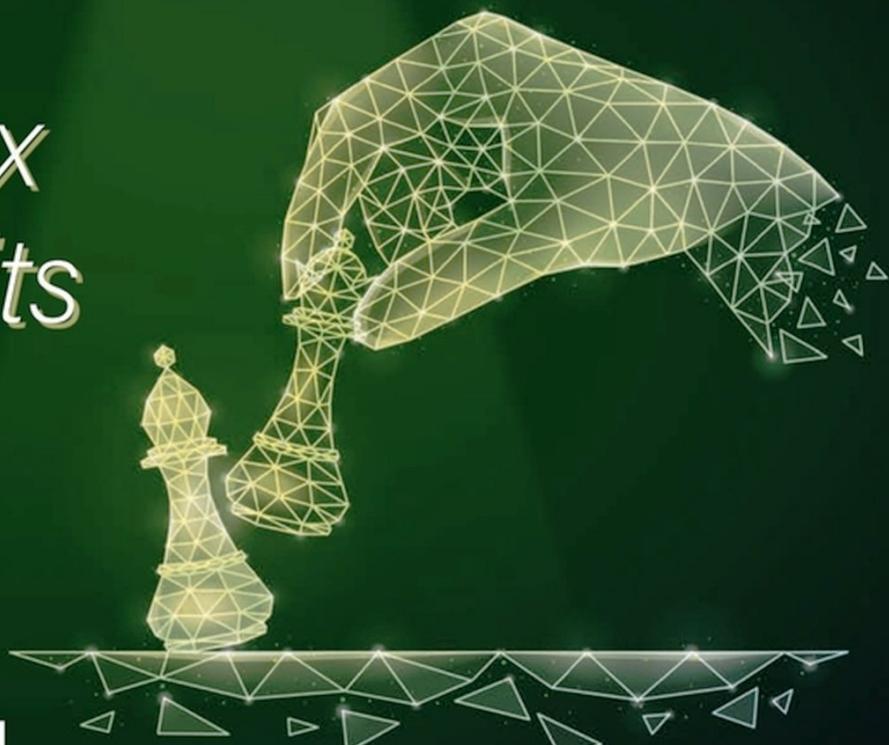
de réduction du temps
d'onboarding

EasyVista aide les entreprises à adopter une approche plus proactive et prédictive en exploitant la puissance de **solutions ITSM et ITOM** !



WWW.EASYVISTA.COM/FR

Les jeux sont faits



par Bertrand Garé

Il y a quelques jours se tenait à San Jose en Californie la première conférence GTC de Nvidia depuis la pandémie. Des milliers de personnes ont participé à cette conférence pour connaître les annonces du fondeur sur ces nouvelles puces graphiques, une technologie que la société domine largement avec près de 80 % de parts de marché laissant loin derrière AMD et Intel. Le fait marquant de cette conférence a été le nombre de partenariats renouvelés, étendus ou annoncés lors de cette conférence. Ainsi NetApp a étendu sa collaboration avec le fondeur. Cloudera va utiliser les micro-services d'intelligence générative de Nvidia. Oracle étend son alliance pour proposer une intelligence artificielle souveraine s'appuyant sur les nouvelles puces Blackwell. Dans le domaine de la sécurité Crowdstrike va proposer les services de Nvidia en IA sur sa plate-forme Falcon. Checkpoint va intégrer les DPU de Nvidia dans sa solution Cloud AI Protect. DDN va faire de même avec les DPU Bluefield-3 dans ses serveurs Exascaler. Snowflake a annoncé aussi l'extension de son partenariat afin d'offrir son data cloud comme backend aux entreprises souhaitant utiliser les technologies de la société de San Jose pour leur analyse en intelligence artificielle. Dell, SAP, Cohesity, Pure Storage, Ansys et même Siemens ont eux aussi fait part de leurs liens avec le fondeur.

Surfant sur la vague de l'intelligence artificielle générative, Nvidia prend donc une place prééminente et explique que le CA de NVIDIA atteint 60,92 milliards de dollars en 2024, soit une augmentation de +125,85 % par rapport à la période précédente. Sa marge brute est aussi passée de 56,9 % en 2023 à 72,7 % en 2024. La capitalisation boursière de NVIDIA est de 2188,2 milliards de dollars en

mars 2024. Elle constitue donc la troisième capitalisation boursière mondiale derrière Microsoft (3.087 milliards) et Apple (2.774 milliards). Mais plus que la domination, parfois jugée outrance du fondeur, il faut y voir, à l'appui des partenariats annoncés, le choix de l'industrie pour les GPU dans les processus et cycle de l'intelligence artificielle.

Il ne reste pas beaucoup de lettres

Cette domination pose aussi question autour des technologies développées par d'autres fondeurs qui ont spécialisé leur puces sur le traitement des données. Quasiment toutes les lettres de l'alphabet sont désormais prises et il ne reste que ZPU et quelques autres pour qualifier ces puces. La plus à la mode avec Groq est le LPU ou Language Processing Unit. Cette dernière société affirme produire des puces dédiées à l'IA offrant une plus grande puissance de calcul que celles de Nvidia. Pour arriver à ses fins, l'entreprise a développé une nouvelle architecture de puce nommée LPU (Language Processing Unit). Gravé en 14 nanomètres, ce composant a pour seul et unique objectif de faire tourner les différents modèles de langages d'intelligence artificielle, dont l'emblématique ChatGPT. On peut aussi citer le TPU de Google ou Tensor Processing Unit. Broadcom a lui pris le x avec ses XPU. Kalray s'est emparé du D avec ses DPU ou Data Processing Unit. L'entreprise a récemment signé un partenariat avec Arm permettant à cet environnement d'avoir une puce spécifique pour le traitement des données en volume. De plus Kalray vient d'annoncer une nouvelle carte d'accélération embarquant quatre de ses DPU pour cbler les marchés des applications de vision intelligente

et d'indexation de données. Mais avec ses 25,8 M€ de chiffre d'affaires la société est loin de Nvidia. Nebulon réalise la même chose mais a choisi le S pour ses SPU ou Service Processing Unit. Le N est la lettre la plus partagée. Plusieurs entreprises développent ainsi des NPU ou Neural Processing Unit comme AMD, Qualcomm, Apple ou Intel. Il ne faut pas les confondre avec les Network Processing Unit, des processeurs dédiés à la gestion des interconnexions réseau. Ce marché à lui aussi ses spécialistes. Au F, vous avez les Floating Point Units. Le M se décile en Memory Protection Unit. Ageia a annexé le P pour ses PPU ou cartes de calcul physique. Le R est très spécialisé dans le graphisme avec les Ray Processing Units tout comme le V et les Vision Processing Units. Le S s'accorde au Snippet Processing Unit et IBM se réserve pour l'instant le Q pour ses Qubit Processing Units. Bref toutes ces innovations et spécialisations trouveront certainement leur place mais certainement pas sur le terrain de Nvidia. On peut se demander ce que vont devenir celles-ci devant l'engouement autour des GPU de Nvidia. D'ailleurs comme nous l'ont fait remarquer en privé certains fournisseurs de serveurs ou de baies de stockage. Le marché actuellement se scinde en deux parties : ceux qui peuvent fournir des GPU... et ceux qui n'en ont pas. Comme Nvidia privilégie les larges commandes de ses puces. Il va y avoir des difficultés pour certains dans le domaine du hardware dans les semaines à venir ! Et c'est sans compter sur les possibles ruptures du fait du récent tremblement de terre à Taïwan même si TSMC a assuré que ses sites avaient été faiblement impacté.

Quel futur ?

On peut déjà s'interroger sur l'avenir de l'historique CPU. Avec la multiplication des puces spécialisées sur certains domaines pour les décharger des tâches trop gourmandes, il est fort possible qu'elle ne serve plus qu'à faire tourner le serveur sous-jacent et orchestrer les différentes puces présentes que ce soit pour l'intelligence artificielle, le réseau ou autre. Au pire elle ne servira plus qu'au boot du serveur laissant les autres puces faire leur travail. Ne sera-t-elle plus dévolus qu'au laptop ou autres tour pour les gamers qui eux aussi choisissent le nec plus ultra des cartes graphiques et des GPU pour éviter les lags et avoir une vision de très haute qualité.

Il en est de même pour le traitement en mémoire. Plusieurs start-up se sont lancées sur cette piste comme Hazelcast, qui vient de rendre des résultats prometteurs, ou GridGain, Mythic. AI ou encore Scaleout Software et Fujitsu... Le futur reste cependant incertain sur la possibilité que cette technologie s'impose largement alors que l'industrie subit le tsunami Nvidia.

Le composable, avec la possibilité avec ces différents processeurs de créer des serveurs totalement adaptés à la tâche auxquels ils sont destinés, mérite donc l'attention dans le futur assez proche. Une entreprise comme Liqid qui permet de désagréger les GPU pour apporter la juste et nécessaire portion de GPU pour réaliser des travaux d'IA est une piste intéressante et évite de gaspiller des ressources non utilisées mais chères et parfois rares ! □ **B.G**





RÉVÉLEZ LA VÉRITÉ CACHÉE DE VOTRE RÉSEAU

AVEC

La plateforme NDR avancée de détection des menaces boostée par l'IA.

- Technologie d'IA **collaboratives**
- Transparency et **explicabilité des IA**
- **88 fois moins** de faux positifs
- Réponse intégrée grâce à la **MITRE ATT&CK**
- Détection des menaces **avancées & inconnues**
- Intégration à **l'écosystème existant**



Composants

L'informatique photonique retrouve enfin la lumière

C'est la technologie qui devait succéder au Silicium lorsque la loi de Moore ne sera plus d'actualité. Les composants photoniques ne sont toujours pas parvenus à faire oublier l'âge du silicium. Pourtant, le besoin de puissance des Supercalculateurs et des IA les remet sous le feu des projecteurs.

C'est carrément une image de science-fiction : des ordinateurs qui ne fonctionnent plus à l'électricité, mais à la lumière. La résistance du Silicium empêchant les électrons de se déplacer à la vitesse de la lumière, pourquoi ne pas s'appuyer directement sur les photons et créer une informatique photonique ? C'est désormais une réalité dans le monde des télécoms, et bientôt, des réseaux locaux avec la technologie POL (Passive Optical LAN), mais créer des processeurs 100 % photoniques et a fortiori des ordinateurs optiques représente encore un sacré challenge scientifique. Ce qui constitue un véritable changement de paradigme pour l'informatique est encore loin d'être une réalité en dépit d'atouts potentiels considérables en termes de rapidité et de consommation énergétique. L'explosion des besoins en puissance de calcul pour la simulation haute performance et surtout l'IA pourrait bien pousser cette nouvelle approche hors des laboratoires. Enfin !

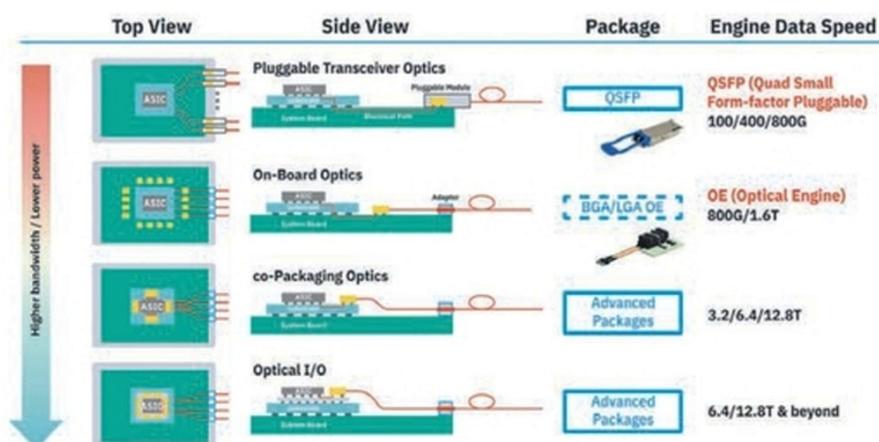
La technologie Silicon-Photonic (SiPh) tient la corde

Depuis 2022, HPE a noué un partenariat avec Ayar Labs. La startup californienne a mis au point un chiplet optique, le TeraPHY, qui offre une bande passante de 4 Tbit/s en



La nouvelle carte Lightelligence portant une puce oNOC conçue pour accélérer les modèles de Machine Learning.

bidirectionnel, pour une consommation de 10 w seulement. Son composant s'appuie sur l'approche SiPh dont l'atout est de coupler une technologie CMOS « classique » à des circuits photoniques à haute vitesse pour fournir une puce peu coûteuse et dans un format très compact. Bien évidemment, la performance de cette hybridation va dépendre de la place prise par le photonique. Pour simplifier, s'il s'agit simplement de placer des transceivers optiques en périphérie du composant électronique, la performance de l'ensemble va être dictée par ce dernier. Pour les composants où la partie optique va jusqu'au cœur de la puce électronique, les performances sont très nettement supérieures.



Tout l'enjeu d'intégrer des technologies photoniques dans des calculateurs traditionnels est de savoir jusqu'où on peut aller sur le volet photonique. Dans cette approche Silicon-Photonic (SiPh), diverses architectures sont possibles, avec un niveau de performance directement en fonction de ce qui est réalisé en photonique et ce qui reste électronique.

Autre startup à suivre, Lightelligence. Cette émanation du MIT travaille depuis 2021 sur le PACE (Photonic Arithmetic Computing Engine), un coprocesseur de calcul optique dont la promesse était d'être 800 fois plus rapide que les GPU du moment. Son architecture baptisée oMAC est, elle aussi, de type SiPh. Flairant l'explosion de la demande de composants pour entraîner les IA, Lightelligence a livré en 2023 une carte intégrant un processeur oNOC, acronyme d'Optical Network-on-Chip (oNOC). L'idée est d'implanter un réseau de 64 noeuds optiques à très haute vitesse et à faible latence pour

accélérer les inférences d'IA. Dévoilée lors de l'été 2023, la carte devait être disponible pour le troisième trimestre avec un SDK permettant à des modèles de Machine Learning d'exploiter sa puissance optique. Avec 64 transmetteurs et 512 récepteurs optiques, cette carte peut être configurée selon plusieurs topologies de réseau optique. Ce sera bien évidemment insuffisant pour déstabiliser NVidia sur le marché des accélérateurs pour l'IA et ses A100 et V100, mais les labos de recherche vont pouvoir commencer à tester cette nouvelle approche. Toutes ces start-up vont clairement devoir relever le défi de faire monter leur technologie à l'échelle et accroître très rapidement la densité et le nombre de nœuds de leurs cartes.

Des innovations de rupture comme s'il en pleuvait

À l'image de la roadmap de ces start-up, la vague d'applications liées à l'IA est bien évidemment ce cas d'usage qui est en train d'insuffler un nouveau souffle d'innovation sur ce domaine. L'empreinte environnementale de l'apprentissage des LLM pose question, ce qui ouvre la voie à des approches alternatives et des innovations de rupture. Or, pour l'instant, les promesses de ces accélérateurs optiques n'ont pas été tenues.

On se souvient qu'HPE avait justement dévoilé un composant 100 % optique en 2017. Financée par la DARPA, cette puce comptait 1052 composants optique et le futur semblait écrit : on allait peu à peu augmenter ce nombre de composants pour en faire un processeur optique surpuissant... un peu comme les processeurs quantiques d'aujourd'hui dont on annonce un nombre de Qubits de plus en plus fantastique. 7 ans après l'annonce, on n'en entend plus parler et le chercheur porteur du projet, Dave Kielpinski, a quitté HPE pour se reconvertis dans... le calcul quantique. Autre candidat en lice avec une approche totalement étonnante, celle du français LightOn, aujourd'hui reconvertis dans les LLM. Son idée était de proposer un OPU (Optical Processing Unit) dont le fonctionnement est pour le moins innovant : Par diffraction multiple de la lumière, le dispositif devait être capable de réaliser des calculs matriciels... à la vitesse de la lumière. L'approche semblait particulièrement adaptée pour implémenter des réseaux de neurones. L'expérience a tourné court et la start-up a pivoté pour exploiter ses connaissances en IA pour proposer des LLM.

Quand l'IA en voit de toutes les couleurs

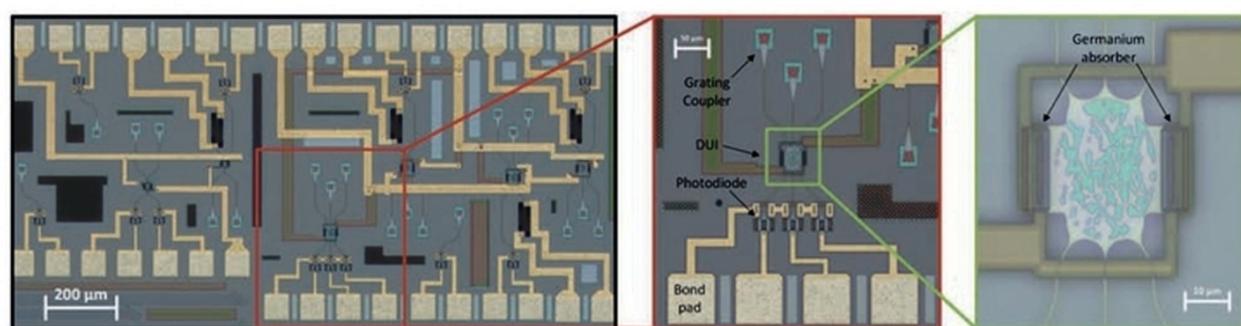
Il y a quelques semaines, d'autres approches particulièrement brillantes ont émergé dans les publications scientifiques. Ainsi, des chercheurs de l'université de Pennsylvanie proposent une nouvelle approche dans la réalisation de composants SiPh. Publiée par *Nature Photonics*, leur recherche vise à exploiter des métastuctures photoniques en silicium de conception inverse afin de réaliser des calculs analogiques avec des ondes électromagnétiques. Les opérations à réaliser sur des matrices de données sont réalisées au niveau des ondes lumineuses directement sur une structure en silicium. Leur recherche montre que l'approche a démontré la pertinence de cette approche sur des matrices 10 x 10.

Une équipe de chercheurs internationaux menés par le Prof. Mario Chemnitz and Dr. Bennet Fischer de l'IPHT de Leibniz à Iéna en Allemagne propose d'utiliser une fibre optique pour simuler les réseaux neuronaux. Les données sont codées sous forme d'impulsions lumineuses ultracourtes de différentes couleurs. Dans la fibre, ces impulsions subissent alors diverses combinaisons, amplifications ou atténuations. Les nouvelles combinaisons de couleurs à la sortie de la fibre permettent de prédire les types de données ou les contextes. Ainsi, un algorithme purement optique est capable de classer les chiffres extraits d'un texte manuscrit : chaque chiffre dispose d'un spectre de couleurs unique, ce qui permet de faire cette reconnaissance à la vitesse de la lumière. L'approche a aussi été appliquée pour la détection des patients atteints de Covid à partir d'un enregistrement audio de leur voix.

Les chercheurs estiment que cette approche est capable d'effectuer diverses tâches couramment entreprises par les réseaux neuronaux artificiels à une fraction seulement de leur complexité d'apprentissage et de leur consommation d'énergie par inférence.

L'électronique sera-t-elle un jour supplantée par la photonique dans le cerveau futur des robots ? C'est possible, mais les enjeux en termes de recherche et surtout de scalabilité de ces approches en rupture sont encore entiers. □

AC



L'intégration du composant optique sur la puce de Silicium telle que l'expérimentent les chercheurs de l'université de Pennsylvanie.

Poste de travail

Un moniteur portable pour un double affichage mobile

Vous en rêviez, Philips l'a fait ! C'est en effet l'une des premières grandes marques à commercialiser un moniteur nomade. Compact, léger et robuste, cet écran Full HD de 15,6 pouces se connecte en un clin d'œil à un PC portable pour profiter d'une solution multi-écran ou partager des présentations partout et à tout moment.



Le socle multi-angle

de l'appareil se révèle très pratique à utiliser au quotidien.

L'inclinaison continue de 0 à 90 degrés permet de le positionner facilement dans n'importe quel endroit.

Al'ère post-Covid, le télétravail à domicile ou en mobilité s'impose un peu plus chaque jour. Cette évolution nécessite toutefois des équipements spécifiques qui soient parfaitement adaptés à ses besoins. Le Philips Business Monitor 16B1P3302D constitue une solution à la fois pratique et performante pour améliorer l'efficacité et le confort de travail, quel que soit l'endroit où l'on se trouve. Encore relativement rare sur le marché, ce type de moniteurs portables permet de bénéficier des avantages du multi-écrans en un clin d'œil. Un compagnon idéal pour compléter l'affichage d'un ultraportable en déplacement, par exemple.

Performances et adaptabilité

Le Philips Business Monitor embarque une dalle de 15,6 pouces au format 16:9 avec un taux de rafraîchissement de 75 Hz. Il délivre une résolution Full HD (1920 x 1080 pixels) avec une luminosité de 250 cd/m² et un contraste natif assez faible de 700:1. Grâce à une finition mate et un traitement antireflet efficace, il demeure parfaitement visible, y compris en plein jour ou dans des conditions de luminosité intense. Basé sur la technologie LED IPS, l'écran LCD possède de larges angles de vision (178/178 degrés) offrant la possibilité de le regarder depuis n'importe quelle position. À l'instar

des moniteurs de bureau professionnels de la marque, l'affichage bénéficie de technologies de pointe telles que le mode LowBlue et anti-scintillement chargé de réduire la fatigue oculaire durant les longues sessions de travail.

Pour piloter l'OSD et effectuer les réglages, on trouve trois boutons de commande à l'arrière du socle. Comme pour la plupart des moniteurs, la navigation dans le menu est un peu acrobatique au départ et nécessite un certain temps d'adaptation. Philips a toutefois intégré son interface maison SmartImage qui permet d'optimiser automatiquement le contraste, la saturation et la netteté des images en sélectionnant différentes modes d'affichage préréglées : bureau, photo, film, jeu, économie, etc. Mention spéciale pour le mode bureau parfaitement calibré en usine pour les usages courants : navigation sur le web, bureautique, lecture de vidéos, etc. Pour un écran d'appoint pour travailler, la qualité de l'affichage est globalement satisfaisante avec des images nettes et détaillées. Au registre des innovations intéressantes, figure le mode EasyRead qui transforme le moniteur en une liseuse avec un écran noir et blanc de type « e-Ink » pour une lecture comme sur du papier.

Construction et connectique

Avec seulement 1,03 kg sur la balance, le Business Monitor est un écran nomade facilement transportable. Livré avec une pochette de protection en tissu, il dispose d'un châssis à l'aspect métallique sobre et élégant. L'appareil s'installe le plus facilement possible grâce à un support intégré au dos à la fois compact, robuste et flexible. Ce dernier permet de positionner l'écran sur un angle pouvant varier de 0 à 90 degrés. Positionnée sur la tranche droite du socle, la connectique se compose de deux ports USB-C et une prise Jack 3,5 millimètres. On regrette l'absence d'un port USB-A et d'une entrée vidéo qui aurait permis de multiplier les cas d'usage avec différents types de périphériques. Des pas de vis



Le Philips Business Monitor 16B1P3302D est un moniteur portable de moins de 1 kg qui permet de créer très simplement un environnement de travail à deux écrans.

de fixation VESA à l'arrière offrent la possibilité d'accrocher l'écran à un mur via un support compatible. Le gros point fort du moniteur réside dans sa capacité à fonctionner via un simple câble double USB-C (avec un adaptateur USB-A amovible) fourni. Inutile de s'encombrer d'un bloc d'alimentation et d'avoir à rechercher une prise secteur pour l'utiliser, il suffit de le relier à un port USB-C (compatible Power Delivery) capable de gérer à la fois l'alimentation, et le flux vidéo.

Créer un environnement de travail proche de ce que l'on peut avoir au bureau en déplacement en seulement quelques secondes, voici la promesse du Philips Business Monitor 16B1P3302D. Cet écran nomade jouit d'une bonne qualité de fabrication, avec un socle à l'ergonomie irréprochable. Le constructeur a mis l'accent sur la simplicité et la praticité en proposant une connectique à la pointe de la technologie que cela soit pour la recharge ou l'intégration à une installation à deux écrans. Optimisé pour la bureautique et la navigation sur le web, l'affichage est parfaitement adapté aux besoins des travailleurs nomades. Le traitement antireflet combiné à la dalle mate garantit un grand confort d'utilisation en pleine journée. Certains regretteront peut-être l'absence d'une surface tactile ou encore la possibilité de le faire fonctionner avec un smartphone ou une tablette. Le constructeur a néanmoins réussi à proposer un tarif relativement contenu de 279 €. □

J.C

CARACTÉRISTIQUES TECHNIQUES DU PHILIPS BUSINESS MONITOR 16B1P3302D

- Écran : LCD LED IPS 15,6 pouces 75 Hz (1920 x 1080 pixels)
- Format image : 16/9
- Temps de réponse : 4 ms
- Niveau de contraste : 700 :1
- Angle de visualisation : 170°/170°
- Modes : LowBlue /EasyRed
- Connectique : 2 x USB-C (Power Delivery), 1 x prise casque 3,5 mm
- Autonomie : 7h32 (en utilisation intensive)
- Dimensions (l x l x h) : 35,9 x 23,2 x 11,9 cm
 - Poids : 1,03 kg
 - Tarif : 279 €

Étude PAC

Une année 2023 solide

Le cabinet d'analyste PAC vient de rendre public son étude sur le marché des services en France pour 2023. Dans l'ensemble les résultats sont plutôt bons même si un ralentissement s'est fait jour.

Apriori tous les signaux sont au vert pour les entreprises du secteur des services en informatique. Quasiment tous les acteurs affichent de la croissance organique de leur chiffre d'affaires et améliorent leur marge opérationnelle. Les besoins croissants en digitalisation des entreprises est un fort soutien de la bonne tenue du secteur indique dans un blog Eric Beaudet. La sécurité et le cloud sont les catalyseurs de la croissance.

Un ralentissement au second semestre

Eric Beaudet précise cependant : « Ces solides performances cachent toutefois un net ralentissement de la croissance, pour tous les acteurs, sur la seconde moitié d'année avec même, pour Capgemini, un recul de son chiffre d'affaires au quatrième trimestre (-0,9 % en organique au niveau mondial). Si ce ralentissement était partiellement anticipé, il a néanmoins été plus brutal que prévu, menant plusieurs acteurs à publier des résultats dans le bas de leurs prévisions annuelles. Capgemini enregistre ainsi une croissance de son CA 2023, à taux de change constant, de 4,4 % (vs +4 % à +7 % visés), Atos une croissance organique de +0,4 % (vs +0 % à +2 % visés) et Econocom de +1,2 % (vs +5 % initialement visés) ».

Ce ralentissement est surtout sensible sur certains secteurs ou zones géographiques. Ainsi le ralentissement a été brutal aux USA. A l'inverse notre pays et l'Europe en général résiste bien avec une croissance forte pour les acteurs principalement exposé au marché français comme Neurones (+ 11%) ou OBS (+6,3%).

LES POSITIONS RESTENT STABLES EN EUROPE

Quasiment dans le même temps que son analyse du marché des services en France, PAC a sorti son classement des 15 principales ESN dans la zone EMEA. Accenture et Capgemini occupent les deux premières places et précèdent TCS et IBM. Malgré ses difficultés Atos reste bien placé. Le reste du classement se compose de grandes entreprises du service et du conseil. Six sociétés sont à la porte de ce classement et luttent pour entrer dans ce classement : EY, Wipro, Fujitsu, Orange, KPMG et HCL.



Les 15 premières ESN en EMEA.

Le secteur TMT souffre de la conjoncture et le secteur financier reste prudent du fait de l'environnement de taux incertain. A l'opposé, la défense, le secteur public et la santé continuent d'enregistrer des rythmes d'investissements soutenus.

Des prévisions optimistes

Si les analystes voient un début d'année difficile en continuation du ralentissement connu en 2023, les besoins en digitalisation et les carnets de commande pleins devraient soutenir la tendance qui devrait donc s'améliorer dès ce trimestre. Les croissances attendues sur 2024 demeurent ainsi positives, quoiqu'inférieures à celles enregistrées en 2023 (+3 % à +5 % attendus pour Econocom ; +2 % à +4 % pour Sopra Steria ; +0 % à +3 % pour Capgemini). Ces éléments sont parfaitement en ligne avec les prévisions. Rappelons que PAC table sur une croissance du marché des services IT, en France, de +3,5 % en 2024, en légère décélération par rapport à 2023 (+4,3 %) avec une plus forte croissance attendue sur les services d'applications (+3,8 %) que sur les services d'infrastructures (+2,8 %). □

B.G

Intelligence artificielle

EY Fabernovel publie un manifeste pour une IA responsable

Alors que la traction business liée à l'IA est une évidence, l'étude souhaite se poser la question de la manière dont l'IA pourrait impacter le bien commun. L'IA peut-elle aider la planète, ou nous aider à la protéger ?

n s'appuyant sur le cadre de réflexion du 6^{ème} rapport du GIEC « *Impacts, adaptation et vulnérabilité* », sur les interactions couplées entre les systèmes climatiques, les écosystèmes et la société humaine, cette étude évalue le potentiel de l'IA à accélérer l'atteinte de trois objectifs clés : « *passer de l'action urgente à l'action opportune* », « *s'adapter à et atténuer* », « *conserver et restaurer* ».

Cette étude se veut « *techno réaliste* » et considère pour chaque cas la dette en ressource que constitue une application IA et les potentielles externalités négatives d'un passage à l'échelle. « *Inaugurer le navire, c'est inaugurer le naufrage ! Et le numérique vient avec un lot de risques réels. La mise à l'eau est pourtant nécessaire* » pour paraphraser la feuille de route numérique et données pour la planification Écologique du gouvernement français.

De l'urgence à l'opportunité

Avec une interface adéquate (outil métier, plateforme), l'IA peut permettre l'utilisation de données peu exploitées aujourd'hui pour les enjeux de durabilité. Elle favorise l'établissement de référentiels communs, trans-sectoriels et internationaux, essentiels pour orchestrer des actions coordonnées et efficaces à l'échelle. Par exemple, comment amplifier la démarche des géo-communs de l'IGN

initiée en 2021 ? Elle pourrait permettre à tous les acteurs d'anticiper des mesures à prendre concernant l'aménagement du territoire, évaluer les risques afin d'atteindre le Zéro Artificialisation Nette (ZAN) d'ici 2050 en se reposant sur une carte de l'anthropocène. La gouvernance des masses de données, leur moyen de collecte et leur accessibilité équitable restent les enjeux clés dans la mise en place de ces outils. Ces questions nécessitent une redéfinition des mécanismes de suivi normatif et des cadres de coopération intergouvernementaux.

Adapter et atténuer

L'intelligence artificielle peut être considérée comme un levier essentiel pour aider nos systèmes de distribution, de production et nos infrastructures urbaines et rurales à optimiser la gestion des ressources, réduire les gaspillages et améliorer l'efficacité énergétique, tout en favorisant une inclusion sociale plus large. Son efficacité dépend d'une implémentation ciblée, guidée par une stratégie de développement durable claire et une approche transparente pour éviter de déplacer ou d'exacerber les problèmes environnementaux, tout se jouant sur la réduction de l'impact net et non sur l'augmentation de l'efficacité relative. En concentrant l'IA sur des objectifs précis, comme la réduction des déchets, l'assistance à l'éco-conception ou la maintenance prédictive, nous pouvons non seulement améliorer l'efficacité et la résilience de nos systèmes mais également contribuer à une transition écologique profonde.

Protéger et restaurer

L'alliance récente de l'IA et de la robotique enclenche une nouvelle phase, celle d'actions autonomes et concrètes pour la restauration des écosystèmes. Ces nouvelles opportunités exigent néanmoins une attention particulière, tant les écosystèmes existants sont complexes et les risques d'effets contre-productifs nombreux comme l'afforestation non contrôlée. □



Un extrait du manifeste de EY Fabernovel réalisé avec le soutien de Microsoft.

B.G

Performance

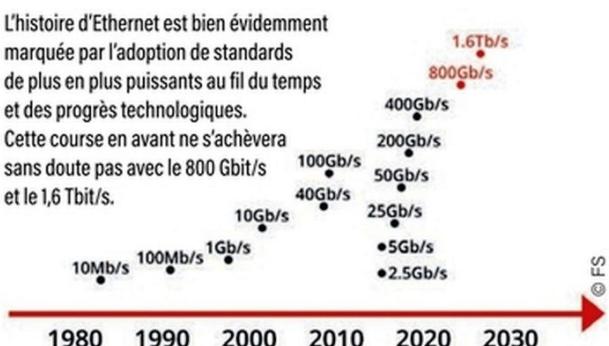
Le Terabit Ethernet dans les starting blocks

Le standard n'est pas attendu avant 2026, et pourtant, les annonces se multiplient dans le Terabit Ethernet. Les composants implémentant le 800 Gbit/s Ethernet et le 1,6 Gbit/s arrivent dans les catalogues, avec un objectif : satisfaire l'appétit insatiable des IA !

En 2023, on fêtait les 50 ans d'Internet, ce réseau inventé par les chercheurs du fameux centre de recherche PARC de Xerox à Palo Alto. Quelques-uns d'entre nous avons connu dans les années 80 ces gros câbles jaunes qui assuraient une connectivité à nos puissants IBM PC au débit stupéfiant de 10 Mbit/s... C'était le fameux Ethernet 10BaseT. Depuis, les débits et les câbles Ethernet n'ont eu de cesse d'évoluer, avec le 100 Mbit/s dans les années 90, puis un réel emballement depuis les années 2000. Ont ainsi déboulé le Gigabit Ethernet, puis le 10 Gigabit, le 40, le 100 jusqu'au 400 Gbit/s. Les réseaux des supercalculateurs sont passés au 400 Gbit/s par agrégation de liens 100 Gbit/s via la technologie SerDes (Serializer/Deserializer), qui n'est finalement qu'une technologie de multiplexage des communications par répartition dans le temps (TDM). En gardant cette approche, mais en passant les liens unitaires à 200 Gbit/s, puis à 400 Gbit/s, le 800 Gbit/s et le 1,6 Tbit/s est désormais à portée.

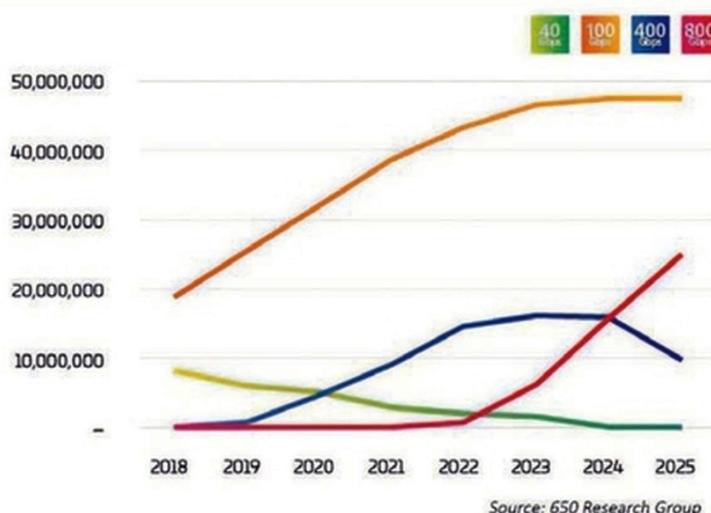
Des fournisseurs de composants qui piaffent d'impatience

Une task force de l'IEEE travaille sur la normalisation du 800 Gbit/s et du 1,6 Tbit/s depuis plusieurs années. Le futur standard 802.3dj comportera plusieurs versions pour des cas d'usage différents. Depuis le 200 Gbit/s exploitant une modulation PAM4 pour les communications chip-to-chip sur paire de cuivre jusqu'au 800 Gbit/s sur fibre monomode avec une portée de 10 à 40 km. L'idée est de normaliser les signaux qui pourront



tant être utilisés pour réaliser des communications en 800 Gbit/s et du 1,6 Tbit/s sur cuivre ou fibre optique, mais aussi du 200 Gbit/s et du 400 Gbit/s sur des liens haute densité et faible consommation. Il s'agira de faire émerger une nouvelle génération de solutions où les utilisateurs pourront utiliser 1, 2, 4 voies d'une solution cuivre ou optique à 8 voies de 1,6 TbE pour supporter un fonctionnement à 200 GbE, 400 GbE ou 800 GbE. On trouvera donc ce Terabit Ethernet dans les switch « Top of Rack », mais aussi entre les rangées de rack et jusqu'à l'interconnexion de serveurs entre datacenters proches, une configuration sur mesure pour les fournisseurs Cloud qui disposent de 3 datacenters par région géographique.

Si les travaux de normalisation du Terabit Ethernet sont planifiés jusqu'en 2027, avec comme point d'orgue la publication du standard prévue pour le mois de juillet 2026, les fabricants de composants ont bien compris que les fournisseurs Cloud et les constructeurs de supercalculateurs ont besoin de plus de débit pour interconnecter les serveurs des clusters et les datacenters entre eux. De facto, depuis deux ans, les annonces et les démonstrations de composants Terabit Ethernet se sont multipliées lors des salons. Dernier en date, le Californien Synopsys qui vient d'annoncer la disponibilité de la première solution IP Ethernet 1,6 Tbit/s du marché. Celle-ci s'appuie sur la couche 224G Ethernet PHY de l'Américain, une modulation PAM4 qui supporte des flux de données de 1,25 à 224 Gbit/s en full-duplex et qui lui permettra de porter des réseaux Ethernet 200G, 400G, 800G, et enfin 1.6T.

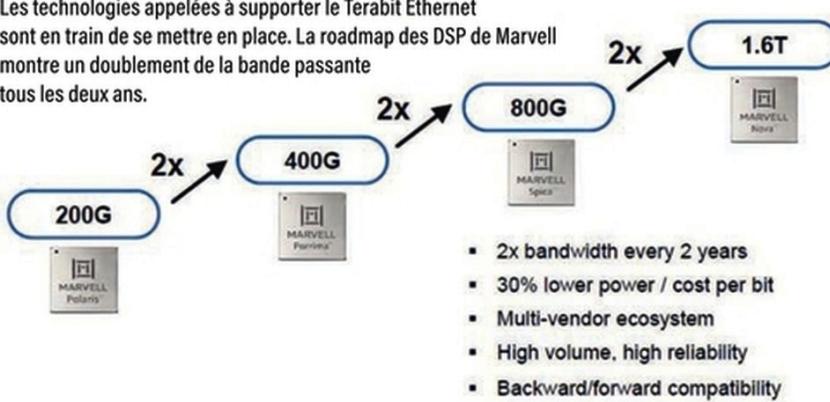


L'explosion des besoins de connectivité causée par le Cloud et de l'IA se traduit très directement par le décollage extrêmement rapide de la livraison de ports 800 Gbit/s dans les datacenters. Le 1,6 Tbit/s devra assurer le relais à l'horizon 2026/2027 avant des débits encore supérieurs.

Selon le fournisseur, la disponibilité d'une solution complète comprenant de nouveaux contrôleurs Ethernet MAC et PCS 1,6T, cette couche PHY Ethernet 224G et une IP de vérification devrait accélérer la mise sur le marché de solutions dédiées au HPC et aux infrastructures IA. Pour l'Américain, cette solution IP Ethernet 1,6 Terabit devrait réduire la consommation électrique des interconnexions en les divisant par 2 comparé aux SoC actuels...

Autre fournisseur très actif sur le front du Terabit Ethernet, Marvell qui a commencé à proposer les premiers échantillons de composants 1,6 Tbt/s dès 2021 avec sa puce Alaska C PHY. De la classe des composants 1,6 Tbit/s Ethernet PHY, le composant agrégeait alors des entrées/sorties 100 Gbit/s PAM4. En mars 2023, lors de la conférence annuelle de l'OFC (Optical Fiber Communication), l'Américain franchissait une nouvelle étape avec son composant Nova. Ce nouveau DSP optique supporte une bande passante de 200 Gbit/s, ce qui constitue une nouvelle étape clé avant l'arrivée de modules 1,6 Tbit/s.

Les technologies appelées à supporter le Terabit Ethernet sont en train de se mettre en place. La roadmap des DSP de Marvell montre un doublement de la bande passante tous les deux ans.



Comme pour Synopsys, outre le doublement de la bande passante, l'accent est mis sur l'efficacité énergétique du composant.

Les fournisseurs de technologies se mettent en œuvre de bataille pour fournir les équipementiers en composants Terabit Ethernet pour que ceux-ci puissent livrer des switchs dès que la norme sera finalisée, sachant que les dernières spécifications techniques auront été finalisées en septembre 2025 avant le vote final en novembre.

Après le Terabit Ethernet, se profile déjà l'Ultra Ethernet

Traditionnellement en avance de phase sur le marché de l'informatique « classique », le monde du HPC travaille déjà sur le futur, sur l'Ultra Ethernet. Créé en juillet 2023 par AMD, Arista, Broadcom, Cisco, Eviden, HPE, Intel, Meta et Microsoft, le consortium Ultra Ethernet Consortium (UEC) est rapidement monté en puissance et a accueilli 27 nouveaux membres en novembre 2023. Celui-ci travaille sur les multiples goulets d'étranglement qui brident les performances des grands clusters de calcul tant sur les couches physiques, réseaux que logicielles et stockage. Ses experts travaillent notamment sur un protocole de communication radicalement nouveau pour remplacer le RDMA (Remote Direct Memory Access over Ethernet). Initialement créé par l'IBTA (InfiniBand Trade Association), ce protocole permet de charger les données directement dans la mémoire de la machine. C'est bien évidemment indispensable pour booster les performances des clusters de calcul et les architectures IA, mais le RDMA se trouve aujourd'hui quelque peu submergé par les volumes de données et la taille des clusters de l'IA qui sont bien au-delà de ce à quoi s'imaginent ses concepteurs.

Le consortium Ultra Ethernet travaille sur un tout nouveau protocole, l'UET (Ultra Ethernet Transport). Parmi les caractéristiques de ce protocole conçu pour fonctionner sur Ethernet et sur IP, celui-ci gérera les trajets multiples sans besoin de répartition de charge, ni contrôleurs de routage, mais par une optimisation par IA pour éviter les congestions réseau. Jusqu'à un million de terminaisons réseau seront gérées. Ses concepteurs estiment que celui-ci est conçu pour tirer profit des équipements 800 Gbits et 1,6 Tbits et ceux qui leur succéderont. Le consortium UEC est hébergé par la fondation Linux et a annoncé une collaboration avec la fondation Open Compute Project qui élaborera les datacenters et équipements de demain. La mise en synergie des travaux des deux organisations devrait sans doute faciliter l'essor de nouveaux standards pour les supercalculateurs et datacenters du futur. □

AC



La fibre et le sans-fil en croissance

Les dernières statistiques de l'OCDE montrent que la fibre optique et l'accès hertzien fixe (AHF ou accès fixe sans fil) ont connu la plus forte croissance des technologies fixes à haut débit depuis trois ans.

Les abonnements à la fibre optique ont augmenté de 56 % entre juin 2020 et juin 2023, et les abonnements à l'AHF ont augmenté de 64 %. Les États-Unis (252 %), l'Estonie (153 %), la Norvège (139 %) et l'Espagne (118 %) ont mené cette croissance de l'AHF. Le dynamisme de la fibre optique et de l'AHF contraste fortement avec le déclin du DSL (- 24 %). La fibre, qui est la technologie dominante depuis 2021, représente 41 % du total des abonnements au haut débit fixe en juin 2023. La technologie AHF représente encore une part plus modeste de 3,7 % du total des abonnements au haut débit fixe (dans les pays où les données sont disponibles). Le DSL voit désormais sa part atteindre 22 % du total du haut débit, tandis que le câble représente 30,5 % de l'ensemble des connexions. Le satellite, qui a progressé de 11 % au cours des trois dernières années, représente maintenant 0,7 % des abonnés au haut débit fixe.

La fibre majoritaire ?

Neuf pays de l'OCDE ont plus de 70 % de connexions par fibre optique par rapport au total du haut débit : la Corée, le Japon, l'Islande et l'Espagne arrivant en tête avec les taux de pénétration de la fibre les plus élevés (89 %, 86 %, 85 % et 84 %, respectivement). C'est en Europe que les taux de croissance de la fibre optique sont les plus élevés, l'Autriche et la Belgique affichant des taux de croissance de 75 % et 73 % au cours de l'année écoulée, suivies de près par le Mexique avec une croissance de la fibre optique de 68 %. Deux autres pays d'Amérique latine figurent dans le top 7 : le Costa Rica et la Colombie, avec des taux de croissance de la fibre optique de 42 % et 34 %, respectivement.

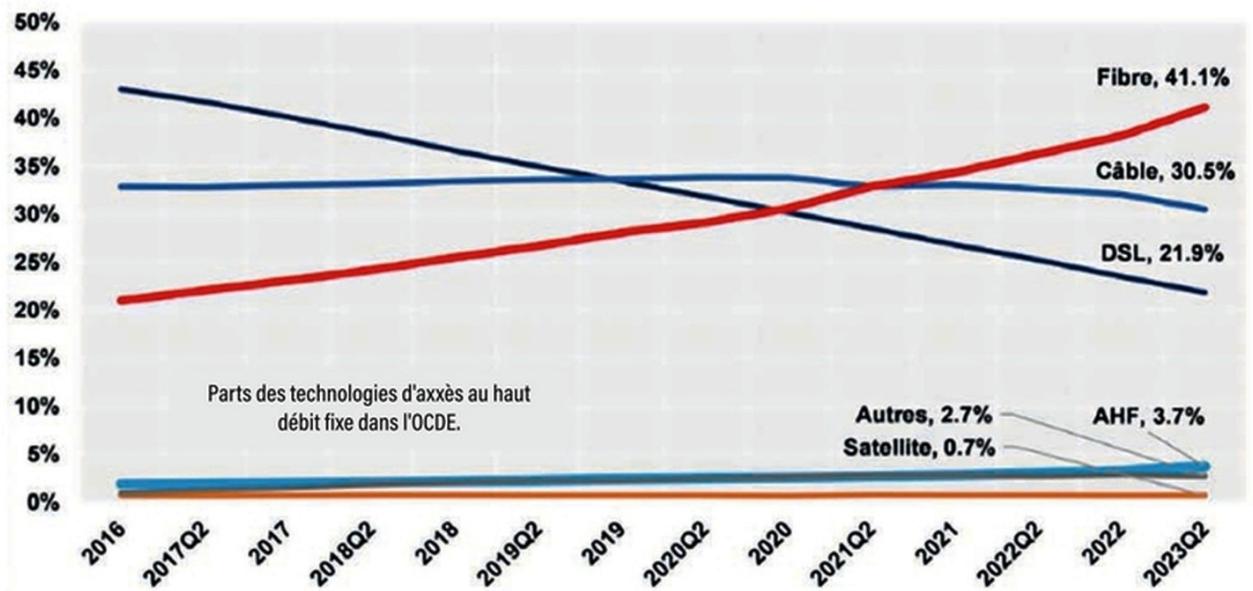
Le mobile support des données

La consommation de données mobiles par abonnement a considérablement augmenté de 28 % en un an, passant de 10,2 Go à 13 Go par abonnement et par mois dans les pays de l'OCDE en juin 2023. La quantité de données consommées dans les pays varie considérablement, allant de 6 Go à 46 Go, la Lettonie étant le leader de l'OCDE.

Malgré un taux de pénétration du haut débit mobile déjà très élevé dans la zone OCDE, le nombre total d'abonnements mobiles continue d'augmenter de 4,6 % au cours de l'année écoulée, pour atteindre 1,8 milliard en juin 2023, contre 1,74 milliard un an plus tôt. La pénétration du haut débit mobile est la plus élevée au Japon, en Estonie, aux États-Unis et en Finlande, avec des abonnements pour 100 habitants de 200 %, 192 %, 183 % et 161 %, respectivement. Dix-huit pays ont été en mesure de fournir le nombre de leurs abonnements 5G séparément des abonnements au haut débit mobile. La part de la 5G dans le total des abonnements au haut débit mobile est de 23 % en moyenne pour les pays de l'OCDE qui ont fourni ces données.

Les cartes SIM Machine-to-machine (M2M) ont augmenté de 14 % en un an. Les deux pays leaders sont la Suède avec 238 cartes SIM M2M pour 100 habitants et l'Islande (203), suivis par l'Autriche (179), les Pays-Bas (93) et la Norvège (76). La Suède et l'Islande émettent toutes deux des cartes SIM M2M à usage international. □

B.G



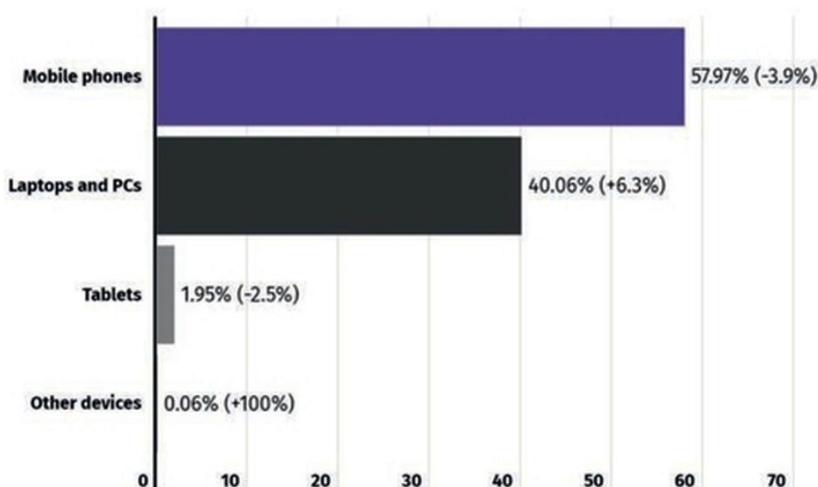
Mobile Le trafic mobile Web en baisse

Selon une étude de Stocklytics.com, le trafic web mobile connaît une seconde baisse historique en 10 ans et représente 57,9 % du trafic Web.

Si 97 % des utilisateurs d'Internet possèdent un smartphone, le trafic en lui-même a baissé de 4 % en 2023 s'éloignant de son pic de 2022 (60,3 %). Ces chiffres sont à mettre dans le contexte particulier de la fin de la période de pandémie de la Covid et du retour au bureau dans la plupart des entreprises. Ainsi, le trafic sur PC et des ordinateurs a, quant à lui, augmenté de 6,3 %. Il représente plus de 40 % du trafic sur le Web. Les tablettes sont elles aussi en perte de vitesse et représentent maintenant moins de 2 % du trafic. A contrario, les consoles de jeux augmentent dans une faible proportion avec 0,06 % du trafic. Leur usage a doublé en un an. Seul l'usage de la simple navigation connaît une croissance selon les chiffres compilés par Stocklytics.com.

Percentage of total web pages served to web browsers running on each kind of device in 2023

Source: DIGITAL 2024 Global Overview Report



Des disparités selon les pays

Ces chiffres moyens ne reflètent que peu certains contextes nationaux. Ainsi, selon des données de DataReportal, le Nigeria est le plus gros consommateur mondial de trafic Web mobile. Celui-ci représente 87,6 % du trafic total dans

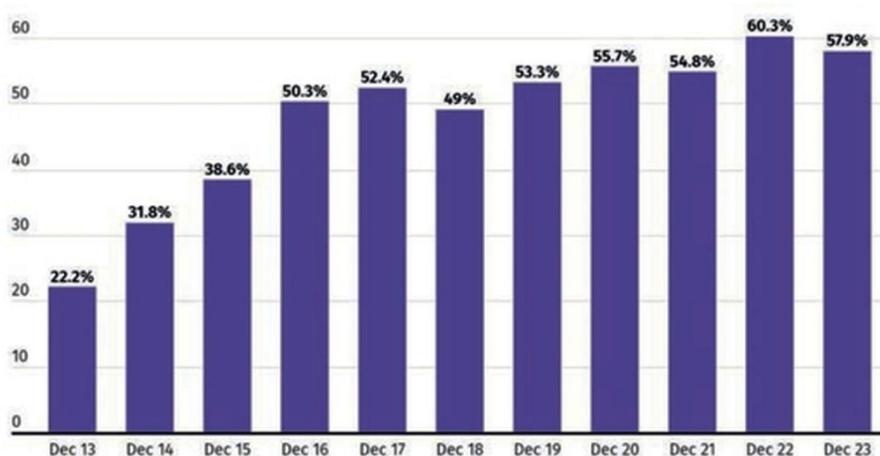
ce pays. Le Vietnam et l'Afrique du Sud suivent avec respectivement 84,5 % et 82 % de leur trafic total.

La plupart des pays occidentaux sont largement en dessous de la moyenne mondiale. L'Allemagne et le Royaume-Uni ont un trafic web mobile de seulement 48,2 % et 47,2 %. Ce ne sont cependant pas les pays où le trafic mobile est le moins grand. Les USA, le Japon ou la Suisse ont des taux extrêmement bas : 41,4 %, 35,6 %, et 35,6 %. C'est cependant le Danemark qui détient la palme du plus petit utilisateur de trafic mobile avec 28,1 %.

Certains de ces chiffres peuvent étonner, car les pays nordiques, le Japon ou les USA sont souvent mis en avant sur leur utilisation de la donnée sur Internet. Il semblerait donc que la réalité ne vole pas au secours des volontés marketing des opérateurs ou des autres acteurs du secteur de l'Internet qui ont pour fonds de commerce d'expliquer que notre pays est en retard sur les usages avec les pays précités. □

Mobile's share of web traffic, from December 2013 to December 2023 (in percentage)

Source: DIGITAL 2024 Global Overview Report



B.G

Embarqué

LoRa, le nouveau réseau gagnant pour l'IoT

La connectivité LoRaWAN fait partie intégrante de l'écosystème IoT. Nous examinerons dans cet article les spécificités et les atouts de cette technologie de transmission ainsi que ses applications spécifiques et pourquoi elle est tant plébiscitée.

Le protocole LoRaWAN a été créé par des ingénieurs français de la start-up grenobloise Cycleo. Fondée en 2009, l'entreprise a été rachetée en 2012 par Semtech, le spécialiste américain des semi-conducteurs. LoRaWAN est un protocole, tandis que LoRa fait référence à la couche physique du réseau. Les fabricants de puces LoRa versent des royalties à Semtech, mais le réseau en soi est ouvert. Le protocole LoRaWAN figure dans la liste des réseaux LPWAN (Lower Power Wide Area Network). Ce sont des réseaux à faible consommation énergétique offrant une longue portée de transmission. Le terme LoRaWAN signifie Long Range Wide Area Network ou réseau étendu à longue portée. Il s'agit d'un protocole radio permettant la transmission de paquets de données de petite taille en bas débit. L'émission du signal radio se fait sur un large spectre afin d'offrir une bonne robustesse aux interférences et aux signaux parasites. Cette technologie offre une grande simplicité de communication et permet des échanges de données sur de longues distances. La principale différence entre les connectivités cellulaires de type 4G ou 5G et les réseaux LPWAN comme LoRaWAN réside essentiellement dans la quantité de données transmises. Alors que les réseaux 4 et 5G sont conçus pour transmettre de grandes quantités de données (idéal pour les smartphones), les réseaux LPWAN ne permettent la transmission que de petits paquets de données. Bien moins énergivore, cette technologie est très intéressante pour l'IoT.

La connectivité LoRaWAN est robuste aux interférences et peut transporter des données sur de longues distances (jusqu'à 15 km) avec un débit allant de 0,3 à 50 Kbps. Cette robustesse évite trop de répétitions de transmissions radio et, en conséquence, ne nécessite qu'une très faible énergie au niveau des capteurs. Comme le réseau utilise une bande de fréquence sans licence, le coût est nettement moins élevé que sur le réseau concurrent Sigfox, par exemple, qui lui, est propriétaire. Dans les réseaux publics, cette technologie peut assurer la couverture de quelques millions d'appareils connectés. Elle est également très bien adaptée à la couverture de l'intérieur des bâtiments, sous-sols et caves compris. Son architecture totalement ouverte donne à n'importe quelle entreprise la liberté de créer son propre réseau dès lors qu'elle achète des puces LoRa. Selon les infrastructures, l'usage et l'exploitation de la technologie, le réseau en question peut être aussi bien public que privé. Les bandes de fréquence disponibles et compatibles avec cette technologie réseau varient selon les continents, voir les pays. En Europe et en France, par exemple, la bande de fréquence employée par le réseau est

de 868 MHz. LoRaWAN est en fait déjà présente dans notre vie quotidienne, que ce soit pour transmettre des données prises par des capteurs de géolocalisation ou pour localiser des véhicules ou des objets de valeur. Elle est aussi employée dans la logistique pour effectuer les suivis des conditions de transport comme le contrôle de la chaîne du froid, les services d'éclairages publics connectés, les parkings intelligents, les « smart cities » (villes intelligentes hyper connectées) ainsi que dans le domaine de la smart agriculture. Le réseau LoRa n'est pas adapté à un suivi des données en temps réel, mais bien plutôt à une transmission d'informations en continu et de façon régulière.



Un réseau LoRaWAN open source baptisé The Things Network (<https://www.thethingsnetwork.org/>) est déployé dans 89 pays. Il enregistrait en septembre 2023 1,5 million d'appareils connectés à son infrastructure.

La pénétration du réseau LoRa

Ce réseau mixte, utilisant Internet et certaines fréquences radio a le gros avantage de traverser les bâtiments, sous-sols, caves et autres lieux confinés sans peine, facilitant ainsi grandement la transmission et les échanges de données. Cette caractéristique fait de ce réseau un outil intéressant lorsqu'il s'agit de transmettre des données à partir de biens ou équipements sensibles ou de valeur, stockés dans des zones sécurisées et volontairement difficiles d'accès. En France, le réseau LoRa est en perpétuelle expansion. Il est largement porté par les opérateurs Bouygues et Orange. S'il présente des avantages par rapport à Sigfox d'un point de vue géolocalisation et mode de fonctionnement bidirectionnel, il reste néanmoins moins avantageux pour l'instant pour des applications transfrontalières nécessitant un roaming européen. □

T.T

14 | 15 | 16 MAI 2024 MONACO

Rejoignez la communauté

READY
FOR
IT!

Rendez-vous du 14 au 16 mai à Monaco

pour l'évènement incontournable des acteurs engagés dans la transition et la sécurité numériques.

Une occasion unique pour échanger autour de vos enjeux cyber et IT :



L'impact législatif colossal de la réglementation NIS2 sur les entreprises

L'adoption de l'IA et son intégration par les métiers



L'évolution du rôle du DSI en tant que leader stratégique face aux crises et défis technologique

Rencontrez en One to One une sélection des meilleurs offreurs de solutions

ADISTA • ALL4TEC • ANOMALI • APPOMNI • ARCAD SOFTWARE • AUDENSIEL • AVEPOINT • BITDEFENDER • BLUE FILES • CEGEDIM OUTSOURCING • CHROME OS • CISCO • CLOUDFLARE • COREVIEW • CRADLEPOINT • CROWDSEC • CYBER GURU • CYBERSEL • CYBERWATCH • DATACORE • DELINEA • DIGICERT • EASY VISTA • FLARE • ESET • HUAWEI • HUB ONE • IC CONSULT • IDECSI • KEYFACTOR • KNOWBE4 • LOGRHYTHM • MANAGEENGINE • NETSKOPE • NOZOMI • NTT FRANCE • OPEN TEXT • ORSENNA • OVERSOC • PROOFPOINT • PURPLEMET • QUALYS • RIOT SECURITY • SHAREKEY SWISS AG • SITUATION • SOSAFE • THEGREENBOW • THE QA COMPANY • WAKERS • WIFIRST • WIZ • ZSCALER FRANCE

Liste non exhaustive

Vous avez des projets d'investissement en cours ou à venir ?



Pour vous inscrire, scannez ce QR code !

Les inscriptions sont ouvertes et soumises à validation

Suivez-nous !



www.ready-for-it.com



Ready For IT



RFIT_event

DG CONSULTANTS COMEXPOSIUM

Trailblazer DX

L'IA omniprésente

La conférence développeur de Salesforce qui s'est tenu récemment à San Francisco a largement traité de comment intégrer et mettre en ouvre l'intelligence artificielle générative avec à la clé plusieurs annonces pour rendre tout cela plus simple.

La conférence qui réunit l'ensemble des développeurs de l'écosystème Salesforce était cette année monothématique et l'IA générative était dans l'ensemble des conversations et des interventions sur scène. Le principal thème était de remettre l'humain à la barre et d'apporter une expérience différente où l'humain travaille en partenariat proche avec l'intelligence artificielle. D'ailleurs la vision de Salesforce est de voir dans le futur l'humain plus comme un orchestrateur que comme un opérateur.

La concrétisation de cette vision est Einstein Copilot qui sort en version Bêta. Le logiciel se présente comme un nouvel assistant d'IA générative personnalisable et conversationnel pour le CRM. Contrairement à d'autres assistants ou copilotes d'IA qui ne disposent pas de données d'entreprise adéquates pour générer des réponses utiles, Einstein Copilot permet aux clients de Salesforce de générer des réponses en utilisant leurs propres données privées et fiables, tout en maintenant une gouvernance stricte des données et sans nécessiter de formation coûteuse sur un modèle d'IA. Ainsi, Einstein Copilot peut répondre aux questions, résumer et créer de nouveaux contenus, interpréter des conversations complexes et automatiser des tâches de façon dynamique.

Cette innovation est rendue possible grâce à la combinaison d'une interface utilisateur conversationnelle, un grand modèle de langage fondamental et des données d'entreprise fiables. L'intégration de ces trois couches permettent aux utilisateurs de Salesforce d'exploiter la puissance de

l'IA générative et d'interagir avec leurs applications de manière entièrement nouvelle.

Un studio pour personnaliser Einstein Copilot

Avec Einstein 1 Studio les administrateurs et développeurs Salesforce peuvent personnaliser les actions proposées par l'IA, créer et intégrer des instructions sur mesure, et exécuter n'importe quel modèle d'IA, aussi bien sur Einstein Copilot que sur les autres applications CRM de la plate-forme Einstein 1. Le studio se compose de trois modules. Copilot Builder est conçu pour créer des actions personnalisées afin de réaliser des tâches spécifiques. Les administrateurs ou développeurs peuvent utiliser les logiciels ou plates-formes de leur choix comme Apex, Flow ou les APIs de Mulesoft ou des composants comme des prompts pour apporter la possibilité à Einstein Copilot de réaliser certaines tâches dans les workflows de travail. Prompt Builder permet de créer et d'activer des instructions personnalisées dans le flux de travail. Ce logiciel autorise la création et la réutilisation de prompts d'intelligence artificielle générative sans nécessité de coder afin de simplifier des processus complexes. Model Builder permet de créer ou d'importer une variété de modèles d'IA.

Une nouvelle version de Data Cloud

En s'appuyant sur l'architecture de métadonnées de Data Cloud, la nouvelle version offre aux entreprises de nouvelles possibilités d'exploiter leurs données. Data Cloud permet aux entreprises de connecter leurs données où qu'elles soient — dans les applications CRM Salesforce et les fournisseurs de cloud (Amazon Bedrock, Amazon SageMaker, Anthropic, Cohere, Databricks, Google Cloud's Vertex AI, et OpenAI). De plus il est possible de rassembler les données dans une vue unique, et les exploiter grâce à la puissance de l'intelligence artificielle de Salesforce (Einstein) ou des autres LLM du marché. Cette dernière fonctionnalité, appelée Model Builder, permet aux clients de choisir le LLM de leur choix ou créer un modèle d'IA correspondant spécifiquement à une



tâche à effectuer. Model Builder est une solution à la fois « *no code*, *low-code* et *pro-code* » permettant aux entreprises de créer leurs propres modèles d'IA prédictifs formés sur leurs données Data Cloud.

Des garde-fous et de la sécurité

Avec Trust Layer, un ensemble de fonctions qui assurent la sécurité et la gouvernance dans l'utilisation de l'intelligence artificielle générative, les entreprises peuvent maintenant configurer plus facilement le masquage des données. De plus une traçabilité d'audit et des retours sur les données collectées par l'IA dans les prompts et les réponses émises sont désormais stockées dans Data Cloud avec la possibilité de créer des rapports ou de lancer des alertes en cas de violations des règles d'utilisation des données dans l'entreprise au travers de Flow ou des autres outils présents dans Einstein 1.

Le retour d'Heroku

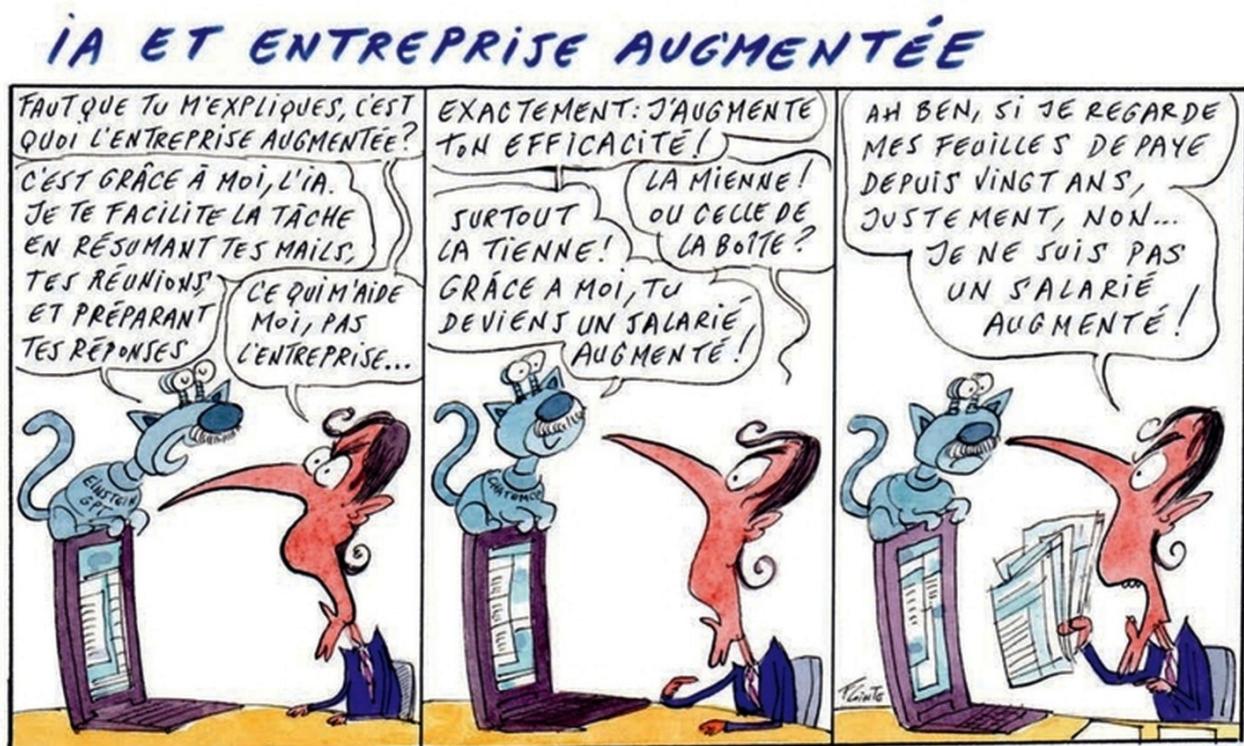
Il y avait longtemps que l'on avait plus entendu parler du PaaS de Salesforce lors des différentes conférences de l'éditeur. Les besoins des entreprises sur l'intelligence artificielle favorisent son grand retour. Elles ont souvent besoin de coder de manière professionnelle quand elles s'engagent dans ce type de démarche. Mais leurs équipes de développement sont souvent séparées ce qui obère leur productivité. Heroku propose de rassembler cela



Parker Harris à son entrée lors de la session plénière de Trailblazer DX.

sur sa plate-forme afin de standardiser la manière de développer des applications dans l'entreprise. Elles vont aussi bénéficier du travail effectué sur la plate-forme qui prend en charge les environnements cloud natifs comme Kubernetes dans Salesforce. La plate-forme continue d'ailleurs son travail d'intégration avec les outils de CI/CD pour rehausser la qualité des développements souvent réalisés par une armée de programmeurs juniors. Bob Wise, le CEO d'Heroku indique : « la moitié du travail est une affaire de base de données. Nous fournissons les services de base de données et la containerisation est masquée sous le capot et nous couvrons les dépôts Github. Tout cela est présent dans Heroku ». □

B.G



Adobe Summit

L'IA pour la création et la collaboration

Au cours de l'édition 2024 d'Adobe Summit qui s'est tenue du 25 au 28 mars à Las Vegas, l'entreprise de San Jose a présenté de nouveaux outils utilisant l'intelligence artificielle (et l'IA générative, bien sûr) pour faciliter la création et la collaboration entre les équipes marketing et les différents partenaires des entreprises. Durant tout le Summit, Adobe a mis en avant ses solutions Adobe Experience Platform ou Adobe Firefly.

Quelque 11000 personnes se sont retrouvées à Las Vegas pour assister au Adobe Summit 2024 (25-28 mars). L'intelligence artificielle était au cœur de la plupart des annonces de l'entreprise avec des solutions comme Adobe Experience Platform ou Adobe Firefly. Au cours de la première journée, l'accent a été mis sur Adobe Experience Cloud, l'environnement qui permet aux marques d'unifier les données. Dans cette suite, Adobe Experience Platform vient apporter une vision des clients sur l'ensemble des canaux avec l'intelligence artificielle qui analyse les données et fournit des informations exploitables en temps réel. Cette fonction est rendue possible grâce à l'introduction du nouvel assistant AI qui apporte une interface conversationnelle capable de répondre aux questions techniques et de simuler des résultats, d'automatiser des tâches et de générer des parcours et des audiences dans les applications Adobe Experience Cloud.

Il faut aussi souligner que l'assistant AI peut simuler des résultats et générer de nouveaux publics et parcours au sein des applications Adobe Experience Cloud telles qu'Adobe Real-Time CDP, Adobe Journey Optimizer et Adobe Customer Journey Analytics. « *Les innovations d'Adobe Experience Platform, telles qu'un assistant AI alimenté par l'IA générative, permettent aux marques de*

comprendre le contexte individuel des clients et de générer des expériences omnicanales sur mesure. Cela augmente la productivité et stimule l'expérimentation de nouvelles idées, tout en garantissant que les marques adhèrent à des contrôles de gouvernance et de sécurité stricts », explique Anjul Bhambri, vice-président senior d'Adobe Experience Cloud.

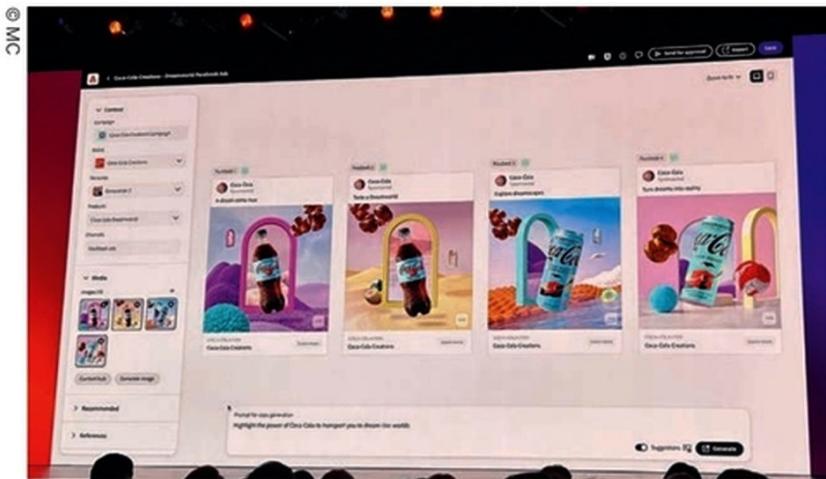
Partage de données entre clients et partenaires

Autre nouveauté : Adobe Real-Time Customer Data Platform (CDP) Collaboration. Avec cette offre, Adobe propose une solution pour gérer les données des clients afin de les utiliser au mieux, tout en collaborant en toute sécurité avec des partenaires, en respectant la confidentialité de ces données. Il s'agit surtout d'aider les marques pour qu'elles puissent découvrir, toucher et mesurer les publics à forte valeur ajoutée, et sans avoir recours à des cookies tiers. Pour favoriser la collaboration entre les différents acteurs, Adobe présente une approche flexible, permettant aux clients d'exploiter en toute sécurité des données provenant de n'importe quelle source, notamment AWS, Databricks, Google, Microsoft Azure et Snowflake, entre autres. Par ailleurs, des entreprises telles qu'Axiom, Blockgraph, Experian, LiveRamp, Merkle et Unified ID 2.0 peuvent aider les marques à élargir leur audience.



Shantanu Narayen, président-directeur général d'Adobe, a ouvert Adobe Summit en évoquant l'accroissement de l'utilisation de l'IA dans la création numérique et dans la collaboration en entreprise.

Toujours dans Adobe Real-Time CDP, mais aussi dans Adobe Journey Optimizer, l'entreprise a introduit Federated Audience Composition. Avec cette solution, les clients de ces plateformes peuvent utiliser directement les données de leurs partenaires (AWS Redshift, Azure Synapse, Databricks SQL, BigQuery et Snowflake) pour enrichir les audiences et attributs à forte valeur ajoutée existants, le tout dans un seul et même système. Selon Adobe, cette fonctionnalité apporte la souplesse nécessaire pour répondre aux besoins d'une entreprise, qu'il s'agisse de



Avec le développement de l'IA générative dans ses applications (comme Adobe Firefly), Adobe propose des solutions complètes pour générer du contenu (texte, image, illustration, etc.) afin de mettre au point des campagnes.

micro-segmentation ou de personnalisation en temps réel, avec une activation à faible latence. Federated Audience Composition permet aussi aux clients de décider des données à utiliser et d'éviter de dupliquer inutilement des ensembles de données ou des modèles d'intégration. « Nous avons créé toute notre offre pour qu'elle soit adaptée aux entreprises avec ce côté sécurité et une intégration complète dans le flux de travail », explique Sophie Yannicopoulos, directrice générale d'Adobe France.

Firefly pour changer la création de contenu

Durant Adobe Summit, un large focus a été effectué sur Adobe Firefly avec le lancement d'Adobe Firefly Services et de Custom Models pour changer l'approche de création de contenu avec une personnalisation accrue. Avec Firefly Services, les marques peuvent accéder à plus de vingt API, outils et services créatifs et génératifs qui éliminent les

tâches manuelles et répétitives. Ces fonctionnalités sont intégrées aux applications Adobe Creative Cloud et peuvent également être intégrées dans n'importe quel flux de travail. En matière de personnalisation, les modèles personnalisés issus des services de Firefly peuvent être facilement partagés au sein des entreprises afin de favoriser la cohérence de la marque entre les équipes de création et de marketing. Surtout, les utilisateurs peuvent générer du contenu spécifique aux produits et aux marques avec la possibilité d'intervenir sur les arrière-plans, les objets, les personnages et les sujets.

GenStudio, une offre d'IA générative

Adobe a aussi annoncé le lancement d'Adobe GenStudio, une offre d'IA générative qui permet aux équipes marketing de planifier, créer, gérer, activer et mesurer rapidement le contenu sur une marque. « Chez Adobe, nous souhaitons utiliser l'IA générative comme un copilote afin d'assister les entreprises dans leur métier », souligne Sophie Yannicopoulos. Concrètement, GenStudio s'appuie sur les services partagés et l'architecture des applications Adobe pour la chaîne logistique du contenu Adobe Workfront, Adobe Experience Manager Assets, Adobe Express, Adobe Firefly, and Adobe Customer Journey Analytics.

En parlant d'Adobe Experience Manager (AEM) Assets, l'application permet de diffuser des ressources gérées par une marque dans l'entreprise, chez les partenaires, etc. « Les utilisateurs peuvent collaborer et partager des images, icônes, illustrations, PDF, vidéos grâce à un hub de contenu unifié pour tous les cas d'utilisation de l'entreprise. Le hub de contenu constitue un point de connexion transparent entre Adobe GenStudio et AEM Assets, où tout utilisateur peut rechercher une ressource et créer des variations personnalisées avec Firefly dans Adobe Express », explique-t-on chez Adobe.

Enfin, il faut aussi évoquer Adobe Content Analytics, dont la vocation est d'aider à mesurer et comprendre les performances du contenu généré par l'IA jusqu'au niveau de l'attribut. Ainsi, le système peut analyser et identifier les éléments (couleurs, objets, styles, etc.) qui trouvent un écho auprès des publics cibles. Il s'agit d'une boucle de rétroaction nécessaire pour s'assurer que le contenu généré par l'IA s'aligne également sur les objectifs de l'entreprise. Bref, l'IA générative s'intègre de plus en plus dans les applications Adobe, et ce n'est que le début. □

Michel Chotard

MICROSOFT COPILOT INTÈGRE ADOBE EXPERIENCE CLOUD

Au cours d'Adobe Summit, la société a annoncé un partenariat avec Microsoft, permettant d'utiliser Microsoft Copilot pour Microsoft 365 dans Adobe Experience Cloud. Cette collaboration doit apporter à Microsoft Copilot des informations marketing et des flux de travail issus des applications Adobe Experience Cloud et de Microsoft Dynamics 365, aidant les équipes marketing lors de l'utilisation d'outils tels qu'Outlook, Microsoft Teams et Word. Avec les fonctionnalités d'Adobe Experience Manager intégrées à Copilot pour Microsoft 365, il est possible de créer des images avec l'IA générative Adobe Firefly ou des textes pour des expériences marketing directement dans Word et de les publier sur le web. « Les équipes marketing passent beaucoup de temps à travailler sur des applications Adobe et Microsoft. Ce partenariat constitue une offre unique, rationalisant les tâches quotidiennes en matière de planification, de collaboration et d'exécution des campagnes », a expliqué Amit Ahuja, vice-président senior Digital Experience Business d'Adobe.

SaaS

Workday veut accélérer sa croissance

Le fournisseur de logiciels RH et Finance continue d'annoncer de fortes ambitions après des résultats de bonne facture et une croissance forte dans la zone EMEA.

Comme chaque année après ses résultats, Workday a organisé un rendez-vous avec la presse pour faire le point sur l'année écoulée et celle à venir. Clairement, 2023 a été un très bon cru avec de la croissance sur tous les produits et d'abord sur le cœur historique de Workday, les ressources humaines. Pierre Gousset indique : « *le sujet devient critique avec les déficiences opérationnelles et notre approche par les « skills » est suivie par des investissements en particulier pour les employés de première ligne, sur les sujets de rétention de la main d'œuvre, la gestion et l'optimisation de la productivité et des coûts* ».

Pour la finance, la croissance s'est révélée aussi forte. Le nombre de clients est maintenant de 1900. « *L'ERP finance commence sa transformation en France après avoir tardé comparativement à d'autres pays européens ou aux USA* » précise Hubert Cotté, DG France de Workday.

Des interrogations sur l'IA

Hubert Cotté, en charge de la France pour Workday est assez prudent : « *les clients sont sur un mode exploratoire. Il y a encore beaucoup d'interrogations et de méfiance* ». Cependant, l'architecture spécifique de la plateforme de Workday autorise de profiter assez simplement des bénéfices de l'IA propriétaire de Workday. Sur des modèles de fondation open source, les équipes de développement de l'éditeur construisent des modèles LLM qui ne sont pas basés sur Internet et certifiés par l'éditeur sur les données structurées et non structurées de l'entreprise. Les modèles peuvent

être enrichis avec des données tierces. L'acquisition et l'intégration de HiredScore comme surcouche sur la plateforme de Workday en est un exemple qui étend les fonctionnalités autour de la recherche de candidats pour des postes vacants.

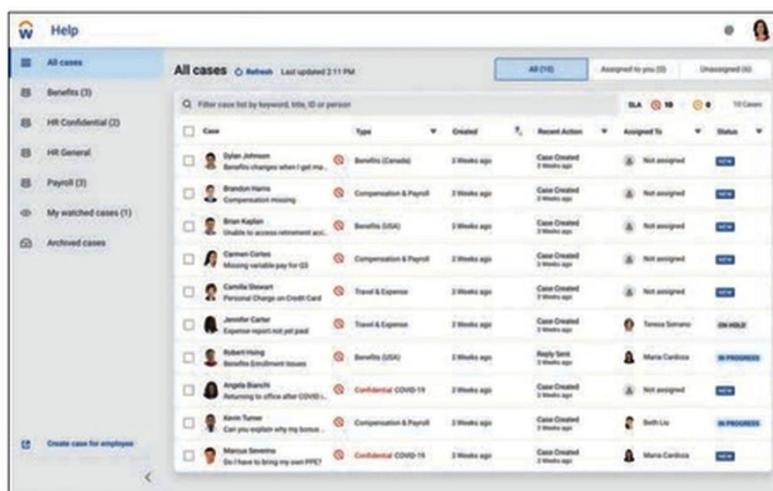
L'IA va surtout infuser peu à peu dans la plateforme avec un co-pilote. Des annonces devraient rapidement arriver lors de la prochaine conférence développeur de l'éditeur avec les fonctionnalités de Copilot dans App Builder, l'outil low code de Workday. Extend, le produit d'extension de la plateforme devrait aussi en profiter avec une proximité plus proche avec App Builder. De la même manière, le module Scheduling de planification va aussi en bénéficier. Hubert Cotté précise : « *les débats, aujourd'hui, sont sur l'organisation et les pratiques de travail dans une phase d'éducation du marché à la recherche de cas d'usages ayant un impact significatif et la possibilité de quantifier cet impact sur le métier de l'entreprise* ».

Des nouveautés à venir

Un premier point à souligner est la montée vers les grands comptes de la solution que ce soit en RH et finance. Workday sort de plus en plus de son identité PME/ETI pour désormais pouvoir fonctionnellement rivaliser avec SAP et Oracle sur ces deux domaines critiques de l'entreprise. Des efforts vont être faits sur l'Entreprise Performance Management et la planification où l'éditeur est confronté principalement à Anaplan et Tagetik. « *Nous devons mieux nous faire connaître dans ce domaine* » assure Hubert Cotté.

De plus, Workday prévoit de développer des versions plus verticalisées et pré-packagées, voire avec des solutions tierces pour répondre encore plus finement aux besoins des clients. L'annonce d'une version Healthcare hors médicaments aux USA pourrait bien être le catalyseur de cette tendance.

Des offres de services, voire de marques blanches pour du business Process Outsourcing, vont certainement se mettre en place. Le partenariat avec ADP autour de FlexSpring pour fournir une solution complète de gestion des ressources humaines, de paie et de conformité en est un premier exemple. □



Un écran de Workday HCM.

B.G

ABONNEZ-VOUS À L'INFORMATICIEN



linformaticien.com/abonnement

MAGAZINE

Recevez chaque mois (10 numéros par an) le magazine «papier» et accédez également aux versions numériques.

1 AN FRANCE : 72 €

2 ANS FRANCE : 135 €

1 AN UE : 90 €

2 ANS UE : 171 €

1 AN HORS UE : 108 €

2 ANS HORS UE : 207 €

NUMÉRIQUE

Accédez chaque mois (10 numéros par an) à la version numérique du magazine et retrouvez également via votre compte en ligne les versions numériques des dernières publications.

1 AN : 49 €

2 ANS : 89 €

ÉTUDIANT / ÉCOLE

Abonnez vos étudiants avec une formule dédiée à 60 % du prix normal de l'abonnement sous forme de PDF (10 numéros par an).

Possibilité abonnements groupés en contactant le service abonnements du magazine à abonnements@linformaticien.com.

ABONNEMENT 1 AN : 43,20 €

Nouvelle version

Oracle présente JAVA 22

Oracle a récemment présenté une nouvelle version de Java qui tire parti de nombreux projets open source. Nous reviendrons plus en détail sur cette version du langage de programmation et de la plateforme de développement numéro un dans le monde.

Sur les multiples améliorations de cette version nous ne reviendrons que sur les principales dans cet article. Celles-ci apportent des améliorations en matière de performances, de stabilité et de sécurité pour aider les développeurs à augmenter leur productivité. Cette nouvelle version apporte notamment des améliorations au langage Java, à ses API et à ses performances, ainsi qu'aux outils inclus dans le Kit de Développement Java (JDK).

La nouvelle version du JDK contient des mises à jour et des améliorations et propose 12 JDK Enhancement Proposals (JEP). JDK 22 apporte des améliorations au langage d'OpenJDK Project Amber (déclarations avant super [...], variables et modèles sans nom, modèles de chaîne, classes et méthodes principales d'instance déclarées implicitement) ; il fournit aussi des améliorations du Project Panama (API de fonction et de mémoire étrangères et API vectorielle), des fonctionnalités liées au Project Loom (concomitance structurelle et valeurs ciblées), des fonctionnalités des bibliothèques et outils de base (API Class-File, lancement de programmes de code source multi-fichiers, et collecteurs de flux) et des mises à jour des performances (épinglage de région pour G1).

Par la canal de Panama

Autre source d'inspiration de cette version, le Projet Panama améliore l'ergonomie, la flexibilité, la sécurité et les performances pour les développeurs en introduisant une API permettant aux programmes Java d'interagir avec du code et des données en dehors de l'environnement d'exécution Java. En appelant efficacement des fonctions étrangères, par exemple, du code en dehors de la Java Virtual Machine et en accédant en toute sécurité à la mémoire étrangère (la mémoire non gérée par la JVM), cette nouvelle API permet aux programmes Java d'appeler des bibliothèques natives et de traiter des données natives sans avoir à recourir à Java Native Interface. De plus une API de vecteur permet aux développeurs d'obtenir des performances supérieures aux calculs scalaires équivalents en introduisant une API pour exprimer des calculs vectoriels qui compilent de manière fiable lors de l'exécution afin de vectoriser les instructions sur les architectures de processeurs (CPU) prises en charge.

Nous reviendrons plus en détail sur cette version dans un prochain article dans notre rubrique DevOps. □ **B.G**

NOUVELLE VERSION DE JAVA



splunk>

Libérez l'innovation avec Splunk.

La plateforme de données pour un monde hybride.

Splunk vous offre une résilience et une sécurité optimales, essentielles pour être innovant. Découvrez pourquoi les plus grandes entreprises du monde entier font confiance à Splunk pour mener leur transformation.

Plus d'infos sur splunk.com/fr

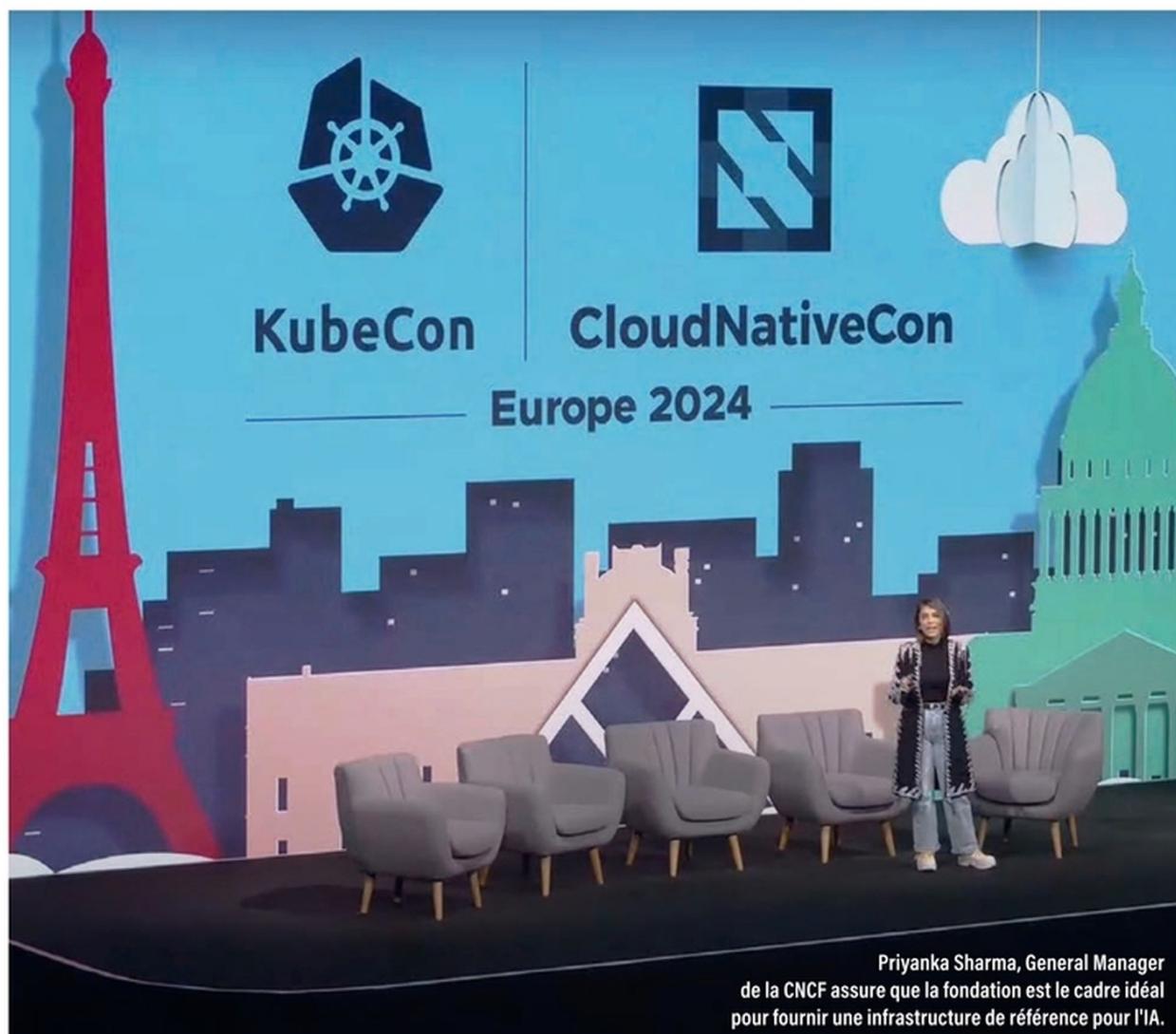


CNCF KubeCon

Kubernetes se rêve en standard de l'IA

Malgré plusieurs écueils inhérents à son fonctionnement, Kubernetes est en bonne voie pour composer les architectures de référence des tâches d'IA générative.

Rassemblé à la KubeCon fin mars, l'ensemble de l'écosystème de la Cloud Native Computing Foundation a, en tout cas, affirmé sa volonté d'avancer dans ce sens.



'IA est définitivement partout. Lors de la KubeCon, événement annuel de la Cloud Native Computing Foundation (CNCF) qui a rassemblé fin mars à la Porte de Versailles l'ensemble de l'écosystème Kubernetes, l'intelligence artificielle générative était sur toutes les lèvres.

Et pour cause, à en croire la plupart des acteurs interrogés sur place, cette dernière ne pourra se faire sans Kubernetes bien qu'il y ait encore quelques problèmes à

résoudre. La couleur a été annoncée dès la keynote d'ouverture où Priyanka Sharma, la General Manager de la CNCF, a rappelé que Kubernetes « contribuait à la création des infrastructures de demain » depuis sa naissance il y a bientôt 10 ans, et donc ne louperait pas le coche de l'IA. « Il y a encore de nombreux défis à relever. Nous avons besoin de standardisation afin d'assurer l'interopérabilité, la sécurité, la gestion des coûts et des ressources, les métriques, ou encore les questions éthiques sur les

traitements IA», a-t-elle ainsi tempéré. Et de son point de vue, la CNCF est le cadre idéal pour y arriver.

Pour illustrer ses propos, ce n'est rien de moins que Timothé Lacroix, cofondateur de Mistral AI qui est monté sur scène aux côtés de Paige Bailey, Lead Product Manager chez Google Deepmind. «Le principal souci concerne aujourd'hui l'allocation des ressources pour les modèles IA», a ainsi expliqué le Français dont l'entreprise opère actuellement plus de 75000 nœuds Kubernetes. Afin d'adresser la question de l'allocation des ressources, Mistral AI s'est tourné vers le projet Slurm pour adresser cette problématique. Ce dernier permet de gérer l'allocation des «nodes» sur Kubernetes pour l'inférence et le fine tuning.

La gestion des ressources au cœur des préoccupations

La question de la gestion des nodes a également fait cogiter chez NVIDIA qui était venue présenter ce qu'ils avaient réalisé dans le cadre du projet de génération d'image Picasso. Le fondateur de Santa Clara expliquait que c'était le principal problème auquel il avait été confronté, avec la tolérance aux pannes et l'optimisation de l'aspect multidimensionnel des modèles. Pour Kevin Klues, ingénieur de la firme qui intervenait sur une session dédiée, la question de la gestion des nodes demande «de changer le mécanisme de règles dans Kubernetes pour demander l'accès à des GPU, et permettre un meilleur partage des ressources au niveau d'un nœud». Pour l'autre grand fondateur, Intel, Patrick Ohly, Cloud Software Architect, avait rapporté dans ses valises le projet DRA (Dynamic Resource Allocation) sur le même sujet. Cette API doit ainsi permettre d'améliorer les capacités de Kubernetes à gérer des ressources locales comme le CPU ou la mémoire et aussi et surtout les GPU.

Autre problème déjà évoqué chez NVIDIA pour assurer le fonctionnement et l'entraînement des larges modèles de données, la question de la tolérance aux pannes semble laisser la plupart des acteurs dans l'expectative. «Cela viendra avec le temps», nous a ainsi simplement déclaré Lachlan Evenson, Principal Program Manager chez Microsoft. Sur ce point, il faut aussi aller chercher des réponses chez les fournisseurs d'infrastructure et de plateforme qui peuvent adresser les problèmes de leur côté. Conscient que l'observabilité jouera d'ailleurs un rôle majeur dans la démocratisation de Kubernetes pour l'intelligence artificielle, OpenTelemetry, le framework open source qui permet aux développeurs de générer, traiter et transmettre des données télémétriques dans un format unique et unifié, a annoncé travailler à la mise en place de nouvelles métriques dédiées à l'IA et aux LLM. Aucune date précise n'a cependant été avancée quant à leur disponibilité.

Pas encore un standard... en l'attente de standards

«Comme d'habitude avec Kubernetes, la question se réglera avec la mise en place de standards que tous pourront utiliser facilement et simplement», résume finalement Shaun O'Meara, le directeur technique de Mirantis. De son point de vue, partagé par Red Hat, l'architecture RAG peut déjà fournir un standard efficace pour l'usage de Kubernetes à des fins d'IA. L'objectif pour les fournisseurs de plateforme va donc être, comme toujours, de proposer des outils pour avoir un usage le plus simple possible de Kubernetes au service de l'IA, avec notamment des initiatives autour de Kubeflow. □

O. Ba

ATTENTION À L'EFFET « BIG DATA »

Quand le Gartner a fait, il y a plus d'une décennie, du Big Data son grand chef de bataille, beaucoup d'entreprises ont commencé à empiler des infrastructures de données colossales et coûteuses sans savoir dans quel but. Et quand les uses case sont enfin arrivées, lesdites infrastructures ne convenaient en fait plus. De nombreux acteurs interrogés lors de la KubeCon ont avoué que cet écueil avait de grandes chances de se reproduire sur l'IA. «Du côté des entreprises, on a une course à ceux qui construiront les plus grosses fermes de GPU alors que les cas d'usages ne sont pas encore arrêtés», explique le responsable d'un grand éditeur Open Source. «Il y a un risque

non négligeable que dans deux ans, les technologies aujourd'hui déployées soient déjà obsolètes et l'argent aura été littéralement jeté par les fenêtres».

Il faut donc aujourd'hui réfléchir à la stratégie IA qui doit être mise en place, d'autant plus qu'il est déjà possible de faire beaucoup sans avoir à investir des sommes colossales. «Que ce soit en travaillant à l'optimisation des LLM Open Source ou des infrastructures, les entreprises peuvent déjà prototyper sur les solutions existantes. Il est tout à fait possible d'avoir de grands modèles d'inférence sur des CPU», explique ainsi Cédric Gégout, Vice-Président en charge du product management chez Canonical. Le cabinet Gartner a

d'ailleurs publié mi-mars une note allant en ce sens. D'après cette dernière, plus de la moitié des entreprises qui ont construit des modèles de langage personnalisés abandonneront ces initiatives en raison des coûts, de la complexité et de la dette technique qu'elles engendrent d'ici 2028. Le cabinet explique que de nouvelles techniques et de nouveaux modèles apparaissent presque quotidiennement pour réduire les coûts et améliorer la précision des réponses. L'analyste Rita Sallam, enjoignait ainsi les entreprises à se tourner vers des architectures découplées, à base d'API, afin d'être en mesure de remplacer rapidement un composant devenu obsolète.

Plateforme

Simplifier l'usage de Kubernetes

Dans sa quête sans fin de rendre Kubernetes plus simple et résilient à utiliser en production et pour des applications critiques, la CNCF doit aussi faire face à l'essor constant de nouveaux cas d'usage et projets. Heureusement, en renforçant les racines, la fondation et ses acteurs arrivent à faciliter le déploiement des ramifications.

vers des environnements Kubernetes toujours plus simples à gérer ? Cheval de bataille de la Cloud Native Computing Foundation (CNCF) depuis la création de l'orchestrateur de container, la simplicité d'utilisation progresse, mais reste encore un axe d'amélioration, d'autant plus que l'ajout constant de nouvelles fonctionnalités ne facilite pas la tâche. « *On arrive aujourd'hui à un certain niveau de maturité pour les utilisateurs qui souhaitent s'appuyer sur Kubernetes comme un socle d'architecture standard pour l'IT* », note David Szegedi, CTO EMEA de Red Hat, qui intervenait lors de la KubeCon qui s'est tenue fin mars à Paris. La généralisation des approches plateformes et managées, à l'inverse du best of breed, y est pour beaucoup.

« *Aujourd'hui, la plupart des entreprises qui viennent nous voir ont des approches Kubernetes-as-a-Service. Notre enjeu est donc d'apporter une plateforme qui est capable de répondre à l'ensemble de leurs usages* », explique David Szegedi. Pour OpenShift, la plateforme de Red Hat basée

sur Kubernetes, cela implique toujours plus d'intégrations et d'opérateurs pour fonctionner, par exemple sur le Edge ou les diverses offres Cloud des acteurs du secteur. Rendre les plateformes Kubernetes le plus agnostique possible est donc un des principaux enjeux de la simplification.

L'orchestration mère de complexité

« *L'idée est d'arriver à des offres de containers en serverless* », note Shaun O'Meara, CTO de Mirantis. « *On s'appuie massivement sur Cluster API pour la gestion des déploiements sur toutes les infrastructures et les standards que nous souhaitons mettre en place avec K0smotron pour la gestion du cycle de vie des control planes tout au long de l'utilisation du cluster* », image le directeur technique qui met également en avant l'usage accru d'API pour pouvoir adresser les autres problématiques liées à l'infrastructure. « *Pour être vertueuse, l'approche plateforme doit répondre à deux principaux enjeux. D'abord être capable de faire fonctionner simplement n'importe quelle distribution de Kubernetes, et assurer une intégration*



optimale avec l'ensemble de l'écosystème de la Cloud Native Computing Foundation », résume Christophe Jauffret, Principal Solution Architect, Nutanix.

Reste que si la base est aujourd'hui simple à utiliser, chaque nouvel usage ou nouvelle brique s'accompagne d'une nouvelle complexité. « *Il faut accepter que nous sommes encore loin d'avoir une solution complète sur étagère* », rappelle ainsi Cédric Gégout, VP Product Management chez Canonical. Pour lui, l'accumulation de couches d'orchestration et d'opérateurs ajoute constamment de la complexité à Kubernetes. « *Il faut aussi que les utilisateurs acceptent que Kubernetes n'est pas la réponse à toutes les problématiques. Nous voyons aujourd'hui des gens essayer de le faire entrer au chausse-pied pour des cas d'usage qui ne sont pas adaptés. Ils créent encore plus de complexité* », ajoute le vice-président. Pour les entreprises souhaitant rester à la pointe des avancées de la CNCF, cela implique d'avoir une politique à deux vitesses. D'une part, d'avoir un socle en production basé sur des versions bénéficiant d'un support à long terme et étant intégrée de manière optimale avec le reste de l'écosystème. D'autre part, d'avoir des solutions upstreams plus expérimentales demandant un investissement technique plus conséquent.

Des data services toujours plus complets

Au-delà des plateformes faisant tourner les containers et Kubernetes, la question du stockage a également apporté son lot de problématiques qu'il a fallu résoudre, notamment en matière de data services. « *Aujourd'hui, il faut accepter que les applications critiques sur Kubernetes ont besoin d'un stockage stateful capable de répondre simplement à l'ensemble des besoins des utilisateurs, y compris en matière de base de données* », tranche Murli Thirumale, General Manager pour Portworx, la plateforme de stockage dédiée à Kubernetes récemment passée dans le giron de Pure Storage. Ainsi, en s'appuyant sur CSI et ses propres opérateurs, la plateforme entend fournir aux clusters de container la même expérience que sur des environnements de VM. « *Nous sommes aujourd'hui capables d'adresser les problématiques de réplication et de déduplication pour Kubernetes sur n'importe quel environnement via la virtualisation des ressources de stockage, tout en gérant automatiquement différentes classes ainsi que les QoS* », explique le dirigeant.

Chez NetApp, les arguments sont les mêmes. « *Nous devons être capables d'avoir les mêmes data services sur Kubernetes que sur des environnements de machines virtuelles traditionnels* », explique Eric Han, VP Product Management du fabricant. Pour y arriver, NetApp a fait le

KUBERNETES AU CHEVET D'OPENSTACK ?

Si Kubernetes s'impose comme un standard d'architecture, cela n'empêche pas d'autres solutions d'exister. Il peut même leur redonner un coup de jeune. C'est le cas d'OpenStack qui, à travers plusieurs projets portés notamment par Mirantis (MOSK) et Canonical (Sunbeam), a gagné en simplicité d'utilisation. « *Le principal problème d'OpenStack réside dans la complexité inhérente à la gestion de ses différentes briques. Or, en intégrant l'ensemble du control plan dans un cluster Kubernetes, on peut complètement effacer cette problématique* », résume Cédric Gégout de Canonical qui équipe aujourd'hui de très nombreux opérateurs télécoms en OpenStack. Chez Mirantis, MOSK (Mirantis OpenStack On Kubernetes), reprend la même idée.

« *Pour les déploiements de Datacenters-as-a-Service, OpenStack reste aujourd'hui la meilleure solution et son intégration à Kubernetes rend enfin possible une automatisation efficace et pérenne des tâches de gestion et de mise à jour* », résume Shaun O'Meara de Mirantis. Pour lui, OpenStack a donc encore de beaux jours devant lui.

choix d'une approche bottom up. « *CSI ne permettant pas de faire tout ce que nous voulions, il a fallu développer tout un set d'API et d'opérateurs capables d'appeler toutes les fonctions de la plateforme de stockage. Le gros enjeu est ensuite de pouvoir orchestrer simplement toutes ces API à partir du control plan* », détaille le Product Manager. Pour lui, l'étape suivante est de pouvoir assurer une mobilité des workloads la plus fluide possible entre les différentes classes de stockage.

Intégrer la gestion du cycle de vie des bases de données

Toujours dans une logique d'avoir des applications stateful, la gestion native des bases de données est également une demande récurrente des utilisateurs Kubernetes. « *Le problème, c'est que Kubernetes ne sait pas comment fonctionne une base de données. C'est un véritable casse-tête de lui faire comprendre comment gérer son cycle de vie* », détaille Kuasi Mensah, Directeur Product Management chez Oracle. L'éditeur a ainsi développé deux contrôleurs Kubernetes pour adresser autant de types de déploiements : d'un côté une base unique et de l'autre, un cluster de base de données. « *L'objectif est de pouvoir automatiser la gestion du cycle de vie des bases de données depuis le control plan du cluster Kubernetes* », précise le dirigeant. D'autres acteurs, aussi bien du côté des bases de données que des infrastructures proposent également des intégrations similaires.

En somme, si l'ensemble de l'écosystème CNCF s'affère à rendre Kubernetes plus simple à utiliser, l'ajout constant de nouveaux cas d'usage, fonctionnalités et outils rend la tâche de plus en plus ardue. « *À chaque fois que nous arrivons à automatiser et à standardiser un aspect de Kubernetes, nous avons trois ou quatre autres nouveaux cas d'usages à prendre en charge. Mais notre mission restera toujours de les rendre le plus simple d'usage possible* », résume ainsi Cédric Gégout. □

O. Ba

Sécurité SCC Enterprise : Google s'attaque à la gestion des risques multicloud

Google Cloud a présenté Security Command Center Enterprise, une solution de gestion et d'atténuation des risques dans les environnements multicloud, qui prend aussi en charge les plateformes concurrentes Amazon Web Services et Microsoft Azure.

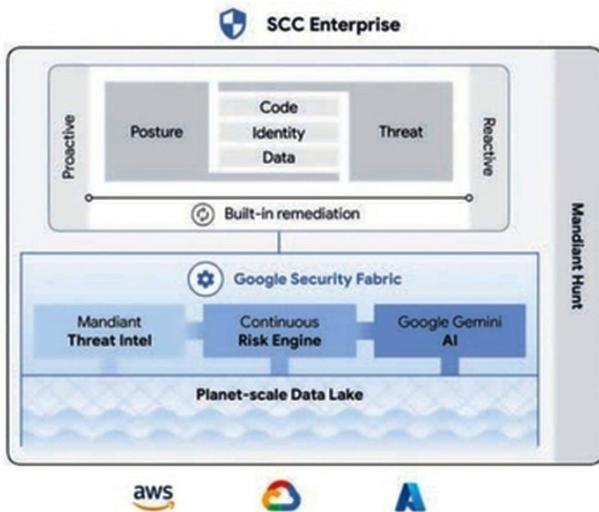
Ce n'est pas une surprise, les organisations adoptent de plus en plus une stratégie multicloud pour héberger leurs applications et leurs données. Cette tendance n'a pas échappé aux APT (Advanced Persistent Threat) qui ciblent de plus en plus ces environnements. Face à ce besoin croissant de sécurisation, les entreprises se bousculent pour lancer sur le marché des plateformes de protection des applications cloud native (CNAPP). Parmi les acteurs majeurs de ce marché, on trouve Wiz, Orca, Microsoft, Palo Alto Networks, Google, et bien d'autres...

Plus de multicloud, moins de silos

C'est au tour de la branche cloud du géant de Mountain View d'annoncer le lancement d'une nouvelle plateforme, Security Command Center Enterprise (SCC Enterprise). Cette solution dédiée à la gestion des risques multicloud combine SecOps alimenté par l'intelligence artificielle et sécurité cloud. Pour se démarquer de la concurrence, la firme mise sur les capacités d'identification des menaces et sur la couverture multicloud plutôt que sur la couverture en silos. Basée sur le Security Command Center, elle s'appuie aussi sur les capacités de renseignements, sur la menace de Mandiant Threat Intelligence, filiale de Google, et sur les capacités d'intelligence artificielle générative de Gemini. La priorité est donnée à la gestion des risques dans les clouds, en intégrant les capacités de renseignement sur la menace de Mandiant afin d'identifier les risques dans les environnements multicloud, alimenter la plateforme de données sur les nouvelles menaces et attaques, et renforcer sa capacité de détection. L'outil peut sécuriser les environnements des concurrents Amazon Web Services et Microsoft Azure, en plus de Google Cloud.

«Notre nouvelle solution peut aider les organisations à imaginer une nouvelle réalité, dans laquelle la visibilité basée sur le SIEM (système de gestion des informations et des événements de sécurité) et la prise d'actions basée sur SOAR (Orchestration, automatisation et réponse aux incidents de sécurité informatique) sont désormais intégrées au monde de la sécurité du cloud», a déclaré Sunil Potti, vice-président/directeur général, de Google Cloud Security.

En pratique, les équipes de sécurité d'une organisation disposent ainsi d'une vue d'ensemble sur les menaces actives, les identités cloud et les données, et peuvent intégrer la



SCC Enterprise peut sécuriser les environnements Amazon Web Services et Microsoft Azure, en plus de Google Cloud.

remédiation et la responsabilité des problèmes dans les flux de travail de bout en bout d'une plateforme cloud. En plus des renseignements sur les menaces de Mandiant, l'outil s'appuie sur un lac de données qui, via l'analyse de données cloud, va générer des graphiques de sécurité permettant de comprendre les interactions à l'œuvre dans l'environnement cloud.

Mandiant en support

Et Gemini dans tout ça ? L'IA générative doit venir en support des utilisateurs les moins expérimentés pour identifier, comprendre, enquêter sur les problèmes de sécurité et les résoudre. Google joue aussi le pentester avec un moteur de risque en mesure de construire un jumeau numérique de l'environnement cloud pour simuler des attaques sophistiquées afin de prédire où les APT pourraient frapper et quelles sources cloud pourraient être exposées.

Pour détecter des menaces qui auraient échappé aux outils de détection traditionnels, la solution fournit une expertise via Mandiant Hunt, une équipe d'analystes et de chercheurs en cybersécurité de Mandiant qui doit venir en support des équipes de sécurité. La firme se risque à dire que la mise à disposition de ces experts pourrait permettre à ses clients de réaliser quelques économies substantielles, en réduisant leur besoin de recrutements. □

V.M



AVANT, LA **DATA**
N'ÉTAIT JAMAIS
ASSOCIÉE À
« **STYLÉ** ».

MAIS ÇA, C'ÉTAIT AVANT
LE PALMARÈS DE
L'INFORMATICIEN.



Qlik France est lauréat dans la **catégorie Business Intelligence** grâce au vote des utilisateurs et à l'accompagnement de nos partenaires. Merci à tous de votre confiance !

Qlik 
TO BE CERTAIN.

Qlik.com

La ville de Mâcon

n'a pas à rougir quand il s'agit de Cyber

Alors que les collectivités sont décriées pour leurs lacunes en matière de cybersécurité, Mâcon fait figure d'impressionnante exception. La ville s'appuie aujourd'hui sur des solutions de pointe alliant, entre autres, protection de l'ensemble des terminaux et des serveurs, segmentation des réseaux et sécurisation des accès.

a Mairie de Mâcon montre l'exemple en matière de compétences cyber. Si les collectivités locales faisaient partie des mauvais élèves au sein du dernier rapport de Cybermalveillance.gouv.fr paru en mars, le directeur général du GIP, Jérôme Notin ayant même largement déploré au téléphone, «*leurs terribles lacunes en matière de cybersécurité*», l'agglomération bourguignonne se démarque avec une approche solide et efficace du domaine. Comptant aujourd'hui 35 000 habitants, la ville a construit au fil des ans une infrastructure solide et en constante évolution, capable de servir efficacement ses agents avec un haut niveau de sécurité. Parmi ses derniers faits d'armes notable, la Mairie a ainsi déployé une solution d'EPDR managée pour l'ensemble de ses agents, instauré une segmentation logique de ses réseaux internes et prépare la mise en place d'un bastion de connexion pour mieux isoler ses systèmes du monde extérieur.

«*C'est au début des années 2000 que nous avons commencé à nous pencher sérieusement sur les questions de cybersécurité. À l'époque, nous avions fait le choix d'équiper tous nos systèmes d'une solution antivirus,*» se souvient Xavier Winckel, Responsable du service systèmes d'informations de la Mairie de Mâcon. Depuis, le service n'a cessé de faire évoluer son infrastructure à un rythme d'une nouvelle solution tous les deux ou trois ans. «*Que ce soit pour la cybersécurité ou les autres pans du SI, nous avons pour politique de faire constamment évoluer nos solutions. Cela nous permet de rester toujours à jour et ça évite aux fournisseurs de se reposer sur leurs lauriers,*» explique le responsable. Menant une veille constante en matière de cyber, le service informatique est ainsi passé par des solutions comme Symantec ou Kaspersky au fil des ans. «*On avait ce dernier en 2022 lors du début de l'invasion Russe en Ukraine, ce qui a accéléré notre volonté de changer. Nous avons donc décidé de nous orienter vers une solution orientée EPDR,*» détaille Xavier Winckel.



Un management granulaire pour ajuster le prix et les performances

Après avoir étudié les différentes solutions disponibles sur le marché, le responsable et ses équipes décident de se tourner vers une solution Watchguard dont il connaissait déjà le savoir-faire via les pare-feux que leur fournissait déjà l'entreprise. «*Même si nous avions historiquement une tradition de gérer toutes nos solutions en interne, nous avons, cette fois-ci, fait le choix de la déployer en mode managé, en partie chez notre partenaire Ava6*», raconte Xavier Winckel. La solution protège ainsi à la fois les serveurs de la Mairie et les terminaux des 400 agents équipés d'outils informatiques. Pour limiter les coûts, la direction a fait le choix de définir précisément quels étaient les terminaux devant bénéficier d'un management poussé de la sécurité. «*Nous n'avons pas besoin d'apporter le même niveau partout. C'est pourquoi nous avons principalement concentré les efforts et les ressources d'Ava6 sur les postes des utilisateurs importants et les serveurs critiques,*» détaille le responsable.

Le projet de sécurisation de la Mairie de Mâcon ne s'est toutefois pas arrêté à cette seule solution. Approchée par l'ANSSI qui souhaitait à la base joindre la communauté d'agglomération, la direction des services IT de la collectivité a pu bénéficier du plan France Relance. Comme quoi les erreurs téléphoniques peuvent avoir du bon. « Nous avons eu droit à un audit qui s'est, pour le coup, plutôt bien passé. Nous n'avons pas eu de mauvaises surprises et les résultats correspondaient à nos attentes », explique Xavier Winckel. Dans le cadre de cet audit, il est ressorti toutefois un besoin accru pour la sécurisation des accès entrants qui doit ainsi aboutir à la mise en place d'un bastion Rubicat. En cours de déploiement grâce à une aide financière de l'ANSSI, il doit permettre d'isoler complètement les systèmes d'information de la Mairie et de mettre en place des connexions sécurisées vers l'extérieur.

Un socle IT solide pour une sécurité plus facile à déployer

« La mise en place du bastion est une continuité logique de ce que nous avions déjà mis en place au niveau de l'infrastructure », raconte le responsable. Après avoir déployé son propre réseau fibré dans toute la ville, la Mairie a centralisé son IT autour de deux salles serveurs basées actuellement sur la solution d'hyper-convergence Simplivity. L'infrastructure réseau est, quant à elle,

LA FORMATION CYBER EST TOUJOURS ESSENTIELLE MAIS...

« Aujourd'hui, ça ne sert à rien de déployer des solutions de sécurité si c'est pour que les utilisateurs fassent n'importe quoi derrière », résume Xavier Winckel. La sensibilisation des différentes équipes sur la cybersécurité est donc indispensable. La Mairie de Mâcon a ainsi engagé un programme de formation continue portant sur la sécurité au sens large. Elle s'est appuyée sur une plateforme d'autoformation achetée via l'UGAP et propose à ses équipes des modules en continu sur les divers aspects de la sécurité.

« Les résultats sont encourageants, car nous savons qui suit quels modules et à quel rythme. Globalement, la plupart des équipes jouent le jeu et nous arrivons à un niveau de sensibilisation satisfaisant », résume Xavier Winckel. Il note toutefois que si la sensibilisation est faite, les comportements restent parfois compliqués à faire évoluer. Même si les populations sont sensibilisées, le responsable remarque que beaucoup continuent à avoir des comportements à risque, même en connaissance de cause. « Il y a évidemment l'aspect pratique qui pousse les utilisateurs à privilier la rapidité à la sécurité, mais aussi un aspect très humain. Plus on va mettre d'interdits, plus les gens vont avoir envie de les dépasser. Dans notre stratégie, nous avons fait le choix d'être le moins contraignant possible avec les équipes, mais d'être intransigeants sur les aspects les plus critiques. Et cela semble fonctionner », explique le responsable.

segmentée logiquement pour que chaque entité ne puisse accéder qu'aux données dont elle a besoin et ainsi éviter que les attaques puissent se propager trop facilement de manière horizontale. Pour Xavier Winckel, la prochaine étape serait de pouvoir déployer un SOC, mais la mise en place de ce dernier reste aujourd'hui conditionnée par l'embauche d'une ressource supplémentaire pour le manager.

Xavier Winckel,
Responsable
du service systèmes
d'informations de
la Mairie de Mâcon.



Conscient que la situation de la collectivité détonne dans un monde où ces dernières sont souvent pointées du doigt pour leurs lacunes en matière de cybersécurité, Xavier Winckel explique cette anomalie simplement : « nous avons eu la chance d'avoir eu très tôt une prise de conscience et une bonne équipe technique. Plutôt que de regarder ce que font nos homologues, nous nous sommes toujours focalisés sur ce que faisaient les entreprises qui sont effectivement plus sensibles que nous sur ces sujets ». Il salue également l'appui des équipes dirigeantes qui, même si elles ne sont pas forcément sensibilisées à la technologie, comprennent les enjeux et abondent dans le bon sens. □

O.Ba

Helpdesk

Arpavie harmonise ses processus avec ServiceNow

Le fournisseur de services à la personne a mis en place un helpdesk en s'appuyant sur les outils de ServiceNow afin d'optimiser son processus. Un projet réussi avec l'idée d'étendre la solution à d'autres sujets.

ARPAVIE propose une gamme complète d'hébergements et de services pour personnes âgées, des plus autonomes aux plus fragiles ou en situation de handicap à travers ses EHPAD, ses maisons de retraite, et son accompagnement à domicile. En 2016, il s'est fait jour qu'il était nécessaire d'harmoniser et de rationaliser les processus liés au Helpdesk. En effet, le courrier électronique était à l'époque, le moyen principal d'échanges entre les utilisateurs des résidences et le support IT pour tout problème informatique. En 2018 il est décidé de mettre en œuvre une démarche de type relation client/fournisseur en interne. Ceci s'est concrétisé par la mise en place d'un centre de services comportant un catalogue de prestations, un centre d'appels, des demandes de maintenance de niveau 1 et 2 ainsi qu'un reporting régulier.

L'idée du projet était ainsi d'inverser le processus : le Groupe fournit des services aux résidences, qui sont ses clients, et non l'inverse. Fallait-il encore trouver l'outil adéquat. Après un appel d'offres ServiceNow a été choisi pour sa large couverture fonctionnelle.

En 2019, la DSI créa le premier portail du groupe, grâce à l'outil ITSM de ServiceNow. Par ce dernier, les utilisateurs peuvent demander une assistance informatique à distance ou de proximité. Les techniciens, quant à eux, disposent d'un tableau de bord pour gérer leurs activités en cours et suivre leur performance. L'inventaire des équipements utilisateurs leur est également accessible. Pour la gouvernance, des indicateurs de suivi sont développés avec Microsoft PowerBI en lien avec ServiceNow.

Ils sont présentés sur un écran de supervision en temps réel ou cumulés, et mis à la disposition des utilisateurs sur l'Intranet. Les résultats ne se font pas attendre 85 % des incidents informatiques sont résolus en une heure et l'indicateur de satisfaction des utilisateurs est de 4/5 en 2023.

Une extension sur d'autres lignes de métiers

En 2021, la DSI réalise que sa démarche pourrait intéresser d'autres services du siège, qui gèrent également des demandes d'assistance de la part des résidences, tels que le service Paie, la direction du patrimoine et des achats, le service du développement RH (formation, gestion des talents). Après une présentation aux membres du comité de gestion du CODIR, le champ d'application de ServiceNow ITSM a été étendu aux services support de ces départements. En mai 2023, le catalogue de services « Patrimoine et achats » a été ouvert, ainsi que les catalogues « Gestion des salaires » et « Gestion des talents » en Novembre.

Un partenaire qui aide à la réussite du projet

Arpavie rend hommage au partenaire intégrateur la société de conseil Noveane-groupe Scalian. Le partenaire a su s'adapter au contexte et aux besoins du groupe, ce qui a permis de déployer le projet ServiceNow dans d'autres services que la DSI. La DSI d'ARPAVIE s'est appuyée sur la qualité des services fournis par Noveane-groupe Scalian. La partenaire a ainsi participé à la mise en place du portail de services apprécié de tous les utilisateurs, de sécuriser les données entre les différentes directions, tout en travaillant en mode agile à distance. Ils réalisent les montées de version annuelles et apporte également de l'assistance, du conseil et du développement ainsi que de la veille technique. De plus sur le premier périmètre du projet tous les objectifs ont été atteints avec des indicateurs de performance qui ont augmenté de 40 % depuis le début du projet. □

B.G



Une des résidences d'Arpavie dans la région parisienne.

RAG et LLM : la protection des données au cœur des modèles d'IA générative

Par François-Christophe Jean, Field Technical Director, Cohesity SEMEA.

À près trois ans de travaux, le Parlement européen a approuvé, mi-mars, le projet de loi de régulation de l'intelligence artificielle (AI Act) visant à assurer plus de transparence quant au fonctionnement des modèles d'IA générative à usage public.

Le potentiel de cette technologie est, en effet, exponentiel, et les entreprises du monde entier ont déjà commencé à l'exploiter. C'est ce que met en évidence une étude réalisée par Censusewide pour Cohesity, fournisseur mondial de solutions de gestion et de sécurisation des données assistées par l'IA; si 86 % des 903 entreprises interrogées utilisent déjà des technologies d'IA générative, les risques liés à la conformité et à la protection des données restent les préoccupations principales pour plus d'un tiers des sondés.

La plupart des utilisateurs ne savent pas encore comment interagir de manière optimale avec les moteurs d'IA. Les solutions disponibles sont encore nouvelles et ne sont pas toutes entièrement développées. Les équipes IT, elles, ne pourront pas faire marche arrière, mais lorsqu'elles considèrent l'acquisition de nouvelles solutions, la priorité doit être mise sur le contrôle des données et l'adoption d'une gouvernance de l'IA qui soit responsable.

Les solutions basées sur l'IA comme Cohesity Gaia, sont souvent, par définition, liées à leur propre environnement et fonctionnent de façon transparente, permettant d'évaluer avec précision les risques d'accès aux données. Cohesity Gaia, l'un des premiers assistants de recherche conversationnels du marché, conjugue les technologies de « Retrieval-Augmented Generation » (RAG) aux grands modèles de langages (LLMs) pour interroger avec précision les bases de données secondaires de l'entreprise



en récupérant des informations provenant de sources multiples, et ainsi, générer des réponses sémantiques, plus complètes et fondées sur un ensemble de données constamment mis à jour, sans besoin de réentraînement du modèle. Pour protéger les informations sensibles et aider les entreprises à se conformer aux exigences réglementaires, l'assistant de recherche s'insère dans l'environnement Cohesity Data Cloud, qui utilise des protocoles d'accès strict et une gestion des priviléges administratifs basée sur les principes de sécurité Zero Trust.

S'il peut être tentant de se laisser séduire par l'apparente facilité de mise en œuvre des modèles RAG, la réalité est que le déploiement d'un tel système à l'échelle d'une entreprise est un défi aux facettes multiples. Il requiert de naviguer entre la complexité des données, les problèmes d'évolutivité, les protocoles de sécurité et les obstacles à l'intégration, tout en garantissant la qualité et l'explicabilité des résultats de recherche obtenus.

Lorsque les entreprises considèrent les différentes options dont elles disposent, investir dans la création de solutions en interne, ou s'associer à des experts dans ce domaine, leur décision ne doit pas être simplement guidée par la question de la technologie. Il s'agit de comprendre l'orientation stratégique de l'entreprise et d'aligner les outils sur cette vision à long terme, en s'entourant de spécialistes qui comprennent les problématiques de gestion et de protection des données, ainsi que la manière dont les modèles RAG peuvent être utilisés pour tirer le maximum d'information de leurs données tout en respectant les exigences de conformité en vigueur.

Choix

Quelles bases de données ?

L'intelligence artificielle étant devenue un domaine très concurrentiel, de nombreuses entreprises explorent différentes pistes afin de prendre en charge, le plus efficacement possible, les tâches qui s'y rapportent. Nous allons voir dans cet article quelles sont les solutions proposées.

Une base de données (BDD) d'IA n'est pas un simple référentiel d'informations, mais plutôt un système dynamique et spécialisé conçu en vue de répondre aux exigences complexes des applications d'IA et de ML. Avec la capacité de stocker, d'organiser et de récupérer des données efficacement, les bases de données d'IA fournissent l'échafaudage sur lequel des modèles d'IA révolutionnaires pourront être construits, affinés et déployés.



© Freepik

Prise en charge de l'IA par une base de données

L'objectif de la prise en charge de l'IA par une base de données est d'accélérer les traitements et d'améliorer la flexibilité des algorithmes employés afin de réduire les coûts. Il s'agit également d'intégrer des modèles de ML (Machine Learning) pour prendre des décisions plus efficaces en s'appuyant sur de grands volumes de données. La clef pour sélectionner une base de données

redis

PLATEFORME SOLUTIONS RESSOURCES COMPAGNIE

ANALYSE DE DONNÉES REDIS

Analysez les clés existantes, ajoutez-en de nouvelles, effectuez des opérations CRUD ou en vrac. Affichez les objets dans un format d'objets JSON joliment imprimé et prenez en charge une navigation conviviale au clavier.

Guide rapide du navigateur →

redisgraph

Graphique

Graph 1

Graph 2

Graph 3

Graph 4

Graph 5

Graph 6

Graph 7

Graph 8

Graph 9

Graph 10

Graph 11

Graph 12

Graph 13

Graph 14

Graph 15

Graph 16

Graph 17

Graph 18

Graph 19

Graph 20

Graph 21

Graph 22

Graph 23

Graph 24

Graph 25

Graph 26

Graph 27

Graph 28

Graph 29

Graph 30

Graph 31

Graph 32

Graph 33

Graph 34

Graph 35

Graph 36

Graph 37

Graph 38

Graph 39

Graph 40

Graph 41

Graph 42

Graph 43

Graph 44

Graph 45

Graph 46

Graph 47

Graph 48

Graph 49

Graph 50

Graph 51

Graph 52

Graph 53

Graph 54

Graph 55

Graph 56

Graph 57

Graph 58

Graph 59

Graph 60

Graph 61

Graph 62

Graph 63

Graph 64

Graph 65

Graph 66

Graph 67

Graph 68

Graph 69

Graph 70

Graph 71

Graph 72

Graph 73

Graph 74

Graph 75

Graph 76

Graph 77

Graph 78

Graph 79

Graph 80

Graph 81

Graph 82

Graph 83

Graph 84

Graph 85

Graph 86

Graph 87

Graph 88

Graph 89

Graph 90

Graph 91

Graph 92

Graph 93

Graph 94

Graph 95

Graph 96

Graph 97

Graph 98

Graph 99

Graph 100

Graph 101

Graph 102

Graph 103

Graph 104

Graph 105

Graph 106

Graph 107

Graph 108

Graph 109

Graph 110

Graph 111

Graph 112

Graph 113

Graph 114

Graph 115

Graph 116

Graph 117

Graph 118

Graph 119

Graph 120

Graph 121

Graph 122

Graph 123

Graph 124

Graph 125

Graph 126

Graph 127

Graph 128

Graph 129

Graph 130

Graph 131

Graph 132

Graph 133

Graph 134

Graph 135

Graph 136

Graph 137

Graph 138

Graph 139

Graph 140

Graph 141

Graph 142

Graph 143

Graph 144

Graph 145

Graph 146

Graph 147

Graph 148

Graph 149

Graph 150

Graph 151

Graph 152

Graph 153

Graph 154

Graph 155

Graph 156

Graph 157

Graph 158

Graph 159

Graph 160

Graph 161

Graph 162

Graph 163

Graph 164

Graph 165

Graph 166

Graph 167

Graph 168

Graph 169

Graph 170

Graph 171

Graph 172

Graph 173

Graph 174

Graph 175

Graph 176

Graph 177

Graph 178

Graph 179

Graph 180

Graph 181

Graph 182

Graph 183

Graph 184

Graph 185

Graph 186

Graph 187

Graph 188

Graph 189

Graph 190

Graph 191

Graph 192

Graph 193

Graph 194

Graph 195

Graph 196

Graph 197

Graph 198

Graph 199

Graph 200

Graph 201

Graph 202

Graph 203

Graph 204

Graph 205

Graph 206

Graph 207

Graph 208

Graph 209

Graph 210

Graph 211

Graph 212

Graph 213

Graph 214

Graph 215

Graph 216

Graph 217

Graph 218

Graph 219

Graph 220

Graph 221

Graph 222

Graph 223

Graph 224

Graph 225

Graph 226

Graph 227

Graph 228

Graph 229

Graph 230

Graph 231

Graph 232

Graph 233

Graph 234

Graph 235

Graph 236

Graph 237

Graph 238

Graph 239

Graph 240

Graph 241

Graph 242

Graph 243

Graph 244

Graph 245

Graph 246

Graph 247

Graph 248

Graph 249

Graph 250

Graph 251

Graph 252

Graph 253

Graph 254

Graph 255

Graph 256

Graph 257

Graph 258

Graph 259

Graph 260

Graph 261

Graph 262

Graph 263

Graph 264

Graph 265

Graph 266

Graph 267

Graph 268

Graph 269

Graph 270

Graph 271

Graph 272

Graph 273

Graph 274

Graph 275

Graph 276

Graph 277

Graph 278

Graph 279

Graph 280

Graph 281

Graph 282

Graph 283

Graph 284

Graph 285

Graph 286

Graph 287

Graph 288

Graph 289

Graph 290

Graph 291

Graph 292

Graph 293

Graph 294

Graph 295

Graph 296

Graph 297

Graph 298

Graph 299

Graph 300

Graph 301

Graph 302

Graph 303

Graph 304

Graph 305

Graph 306

Graph 307

Graph 308

Graph 309

Graph 310

Graph 311

Graph 312

Graph 313

Graph 314

Graph 315

Graph 316

Graph 317

Graph 318

Graph 319

Graph 320

Graph 321

Graph 322

Graph 323

Graph 324

Graph 325

Graph 326

Graph 327

Graph 328

Graph 329

Graph 330

Graph 331

Graph 332

Graph 333

Graph 334

Graph 335

Graph 336

Graph 337

Graph 338

Graph 339

Graph 340

Graph 341

Graph 342

Graph 343

Graph 344

Graph 345

Graph 346

Graph 347

Graph 348

Graph 349

Graph 350

Graph 351

Graph 352

Graph 353

Graph 354

Graph 355

Graph 356

Graph 357

Graph 358

Graph 359

Graph 360

Graph 361

Graph 362

Graph 363

Graph 364

Graph 365

Graph 366

Graph 367

Graph 368

Graph 369

Graph 370

Graph 371

Graph 372

Graph 373

Graph 374

Graph 375

Graph 376

Graph 377

Graph 378

Graph 379

Graph 380

Graph 381

Graph 382

Graph 383

Graph 384

Graph 385

Graph 386

Graph 387

Graph 388

Graph 389

Graph 390

Graph 391

Graph 392

Graph 393

Graph 394

Graph 395

Graph 396

Graph 397

Graph 398

Graph 399

Graph 400

Graph 401

Graph 402

Graph 403

Graph 404

Graph 405

Graph 406

Graph 407

Graph 408

Graph 409

Graph 410

Graph 411

Graph 412

Graph 413

Graph 414

Graph 415

Graph 416

Graph 417

Graph 418

Graph 419

Graph 420

Graph 421

Graph 422

Graph 423

Graph 424

Graph 425

Graph 426

Graph 427

Graph 428

Graph 429

Graph 430

Graph 431

Graph 432

Graph 433

Graph 434

Graph 435

Graph 436

Graph 437

Graph 438

Graph 439

Graph 440

Graph 441

Graph 442

Graph 443

Graph 444

Graph 445

Graph 446

Graph 447

Graph 448

Graph 449

Graph 450

Graph 451

Graph 452

Graph 453

Graph 454

Graph 455

Graph 456

Graph 457

Graph 458

Graph 459

Graph 460

Graph 461

Graph 462

Graph 463

Graph 464

Graph 465

Graph 466

Graph 467

Graph 468

Graph 469

Graph 470

Graph 471

Graph 472

Graph 473

Graph 474

Graph 475

Graph 476

Graph 477

Graph 478

Graph 479

Graph 480

Graph 481

Graph 482

Graph 483

Graph 484

Graph 485

Graph 486

Graph 487

Graph 488

Graph 489

Graph 490

Graph 491

Graph 492

Graph 493

Graph 494

Graph 495

Graph 496

Graph 497

Graph 498

Graph 499

Graph 500

Graph 501

Graph 502

Graph 503

Graph 504

Graph 505

Graph 506

Graph 507

Graph 508

Graph 509

Graph 510

Graph 511

Graph 512

Graph 513

Graph 514

Graph 515

Graph 516

Graph 517

Graph 518

Graph 519

Graph 520

Graph 521

Graph 522

Graph 523

Graph 524

Graph 525

Graph 526

Graph 527

Graph 528

Graph 529

Graph 530

Graph 531

Graph 532

Graph 533

Graph 534

Graph 535

Graph 536

Graph 537

Graph 538

Graph 539

Graph 540

Graph 541

Graph 542

Graph 543

Graph 544

Graph 545

Graph 546

Graph 547

Graph 548

Graph 549

Graph 550

Graph 551

Graph 552

Graph 553

Graph 554

Graph 555

Graph 556

Graph 557

Graph 558

Graph 559

Graph 560

Graph 561

Graph 562

Graph 563

Graph 564

Graph 565

Graph 566

Graph 567

Graph 568

Graph 569

Graph 570

Graph 571

Graph 572

Graph 573

Graph 574

Graph 575

Graph 576

Graph 577

Graph 578

Graph 579

Graph 580

Graph 581

Graph 582

Graph 583

Graph 584

Graph 585

Graph 586

Graph 587

Graph 588

Graph 589

Graph 590

Graph 591

Graph 592

Graph 593

Graph 594

Graph 595

Graph 596

Graph 597

Graph 598

Graph 599

Graph 600

Graph 601

Graph 602

Graph 603

Graph 604

Graph 605

Graph 606

Graph 607

Graph 608

Graph 609

Graph 610

Graph 611

Graph 612

Graph 613

Graph 614

Graph 615

Graph 616

Graph 617

Graph 618

Graph 619

Graph 620

Graph 621

Graph 622

Graph 623

Graph 624

Graph 625

Graph 626

Graph 627

Graph 628

Graph 629

Graph 630

Graph 631

Graph 632

Graph 633

Graph 634

Graph 635

Graph 636

Graph 637

Graph 638

Graph 639

Graph 640

Graph 641

Graph 642

Graph 643

Graph 644

Graph 645

Graph 646

Graph 647

Graph 648

Graph 649

Graph 650

Graph 651

Graph 652

Graph 653

Graph 654

Graph 655

Graph 656

Graph 657

Graph 658

Graph 659

Graph 660

Graph 661

Graph 662

Graph 663

Graph 664

Graph 665

Graph 666

Graph 667

Graph 668

Graph 669

Graph 670

Graph 671

Graph 672

Graph 673

Graph 674

Graph 675

Graph 676

Graph 677

Graph 678

Graph 679

Graph 680

Graph 681

Graph 682

Graph 683

Graph 684

Graph 685

Graph 686

Graph 687

Graph 688

Graph 689

Graph 690

Graph 691

Graph 692

Graph 693

Graph 694

Graph 695

Graph 696

Graph 697

Graph 698

Graph 699

Graph 700

Graph 701

Graph 702

Graph 703

Graph 704

Graph 705

Graph 706

Graph 707

Graph 708

Graph 709

Graph 710

Graph 711

Graph 712

Graph 713

Graph 714

Graph 715

Graph 716

Graph 717

Graph 718

Graph 719

Graph 720

Graph 721

Graph 722

Graph 723

Graph 724

Graph 725

Graph 726

Graph 727

Graph 728

Graph 729

Graph 730

Graph 731

Graph 732

Graph 733

Graph 734

Graph 735

Graph 736

Graph 737

Graph 738

Graph 739

Graph 740

Graph 741

Graph 742

Graph 743

Graph 744

Graph 745

Graph 746

Graph 747

Graph 748

Graph 749

Graph 750

Graph 751

Graph 752

Graph 753

Graph 754

Graph 755

Graph 756

Graph 757

Graph 758

Graph 759

Graph 760

Graph 761

Graph 762

Graph 763

Graph 764

Graph 765

Graph 766

Graph 767

Graph 768

Graph 769

Graph 770

Graph 771

Graph 772

Graph 773

Graph 774

Graph 775

Graph 776

Graph 777

Graph 778

Graph 779

Graph 780

Graph 781

Graph 782

Graph 783

Graph 784

Graph 785

Graph 786

Graph 787

Graph 788

Graph 789

Graph 790

Graph 791

Graph 792

Graph 793

Graph 794

Graph 795

Graph 796

Graph 797

Graph 798

Graph 799

Graph 800

Graph 801

Graph 802

Graph 803

Graph 804

Graph 805

Graph 806

Graph 807

Graph 808

Graph 809

Graph 810

Graph 811

Graph 812

Graph 813

Graph 814

Graph 815

Graph 816

Graph 817

Graph 818

Graph 819

Graph 820

Graph 821

Graph 822

Graph 823

Graph 824

Graph 825

Graph 826

Graph 827

Graph 828

Graph 829

Graph 830

Graph 831

Graph 832

Graph 833

Graph 834

Graph 835

Graph 836

Graph 837

Graph 838

Graph 839

Graph 840

Graph 841

Graph 842

Graph 843

Graph 844

Graph 845

Graph 846

Graph 847

Graph 848

Graph 849

Graph 850

Graph 851

Graph 852

Graph 853

Graph 854

Graph 855

Graph 856

Graph 857

Graph 858

Graph 859

Graph 860

Graph 861

Graph 862

Graph 863

Graph 864

Graph 865

Graph 866

Graph 867

Graph 868

Graph 869

Graph 870

Graph 871

Graph 872

Graph 873

Graph 874

Graph 875

Graph 876

Graph 877

Graph 878

Graph 879

Graph 880

Graph 881

Graph 882

Graph 883

Graph 884

Graph 885

Graph 886

Graph 887

Graph 888

Graph 889

Graph 890

Graph 891

Graph 892

Graph 893

Graph 894

Graph 895

Graph 896

Graph 897

Graph 898

Graph 899

Graph 900

Graph 901

Graph 902

Graph 903

Graph 904

Graph 905

Graph 906

Graph 907

Graph 908

Graph 909

Graph 910

Graph 911

Graph 912

Graph 913

Graph 914

Graph 915

Graph 916

Graph 917

Graph 918

Graph 919

Graph 920

Graph 921

Graph 922

Graph 923

Graph 924

Graph 925

Graph 926

Graph 927

Graph 928

Graph 929

Graph 930

Graph 931

Graph 932

Graph 933

Graph 934

Graph 935

Graph 936

Graph 937

Graph 938

Graph 939

Graph 940

Graph 941

Graph 942

Graph 943

Graph 944

Graph 945

Graph 946

Graph 947

Graph 948

Graph 949

Graph 950

Graph 951

Graph 952

Graph 953

Graph 954

Graph 955

Graph 956

Graph 957

Graph 958

Graph 959

Graph 960

Graph 961

Graph 962

Graph 963

Graph 964

Graph 965

Graph 966

Graph 967

Graph 968

Graph 969

Graph 970

Graph 971

Graph 972

Graph 973

Graph 974

Graph 975

Graph 976

Graph 977

Graph 978

Graph 979

Graph 980

Graph 981

Graph 982

Graph 983

Graph 984

Graph 985

Graph 986

Graph 987

Graph 988

Graph 989

Graph 990

Graph 991

Graph 992

Graph 993

Graph 994

Graph 995

Graph 996

Graph 997

Graph 998

Graph 999

Graph 1000

56 | L'Informaticien n°225 – Avril 2024

d'IA adaptée réside dans l'alignement des fonctionnalités et des capacités de la dite BDD avec les exigences spécifiques d'un projet. C'est donc en évaluant soigneusement des facteurs tels que l'évolutivité, la sécurité, la cohérence des données et la prise en charge de différents types et structures de données que des développeurs pourront faire le bon choix qui contribuera au succès de leurs projets d'IA et de ML.

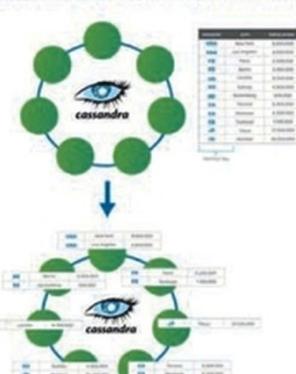
Les grands fournisseurs de BDD

Hadoop est une plateforme open source basée sur Java. Elle fournit des outils pour le traitement du Big Data et propose des bases de données telles que HBase ou MapReduce qui permettent de traiter d'immenses quantités de données. La plateforme propose, en outre, des outils intégrés tels que Submarine pour la prise en charge des algorithmes d'IA et notamment le deep learning. Autre monstre des BDD, Oracle, qui fournit avec sa base plusieurs outils de data science ou de data analytics. Son interpréteur Python intégré permet aux développeurs d'exécuter des algorithmes de ML. Oracle Machine Learning for R permet d'analyser les données avec le langage R. Côté Big Blue, DB2 d'IBM prend en charge de nombreuses tâches propres au ML en langage SQL étendu, comme la régression linéaire. IBM Watson Studio permet, quant à lui, d'accélérer la prise en charge des modèles d'IA en se connectant directement à la base de données. Comprend d'Amazon donne, elle, la possibilité d'exploiter le traitement du langage naturel (NLP pour Natural Language Processing) afin de mieux analyser les données.

STOCKAGE DANS LE CLOUD ET STARTUPS

Les fournisseurs de stockage cloud proposent, eux aussi, des solutions basées sur l'IA. Google Drive, par exemple, fournit des outils tels que AutoML ou Firebase ML permettant de déployer les modèles TensorFlow Lite. Microsoft Azure propose, pour sa part, plusieurs bases de données prenant en charge l'IA telles que Cosmos DB, Data Lake, Neo4j ou encore l'incontournable Elasticsearch. Il existe aussi des startups proposant des solutions de traitement de tâches de type IA/ML. SingleStore vend des logiciels de télémétrie en temps réel, tandis que MindsDB étend le langage SQL afin d'exécuter des routines de ML sur des bases de données classiques. Riskified ou H2O.ai priorisent, quant à elles, le traitement des données, laissant à l'arrière-plan l'aspect base de données.

Want more power? Add more nodes



One reason for Cassandra's popularity is that it enables developers to scale their databases dynamically, using off-the-shelf hardware, with no downtime. You can expand when you need to – and also shrink, if the application requirements suggest that path.

Perhaps you are used to Oracle or MySQL databases. If so, you know that extending them to support more users or storage capacity requires you to add more CPU power, RAM, or faster disks. Each of those costs a significant amount of money. And yet: Eventually you still encounter some ceilings and constraints.

In contrast, Cassandra makes it easy to increase the amount of data it can manage. Because it's based on nodes, Cassandra scales horizontally (aka scale-out), using lower commodity hardware. To double your

Apache Cassandra est une base de données AI NoSQL hautement évolutif privilégiée par de grandes plateformes telles que Instagram ou Netflix

Les meilleures bases de données d'IA

Le choix d'une base de données d'IA appropriée est, comme nous l'avons mentionné plus haut, une considération cruciale pouvant avoir un impact très significatif sur le succès des projets. Les diverses options de bases de données disponibles offrent une large gamme d'options, chacune adaptée pour répondre à des exigences et à des préférences bien spécifiques.

Amazone Aurora

Amazone Aurora offre une compatibilité avec MySQL et PostgreSQL. Sa capacité à évoluer de manière transparente, ses fonctionnalités de sécurité robustes et ses mécanismes de sauvegarde automatisés la positionnent comme une option intéressante pour les applications d'IA et de ML. Les organisations utilisant Amazon Aurora bénéficient de sa gestion efficace des données complexes.

Apache Cassandra

Apache Cassandra (<https://cassandra.apache.org/>) est devenue une base de données NoSQL hautement évolutif. Son architecture décentralisée et son partitionnement automatique lui permettent de gérer efficacement de très grandes quantités de données. Cela la rend particulièrement adaptée aux applications nécessitant des niveaux d'évolutivité et de tolérance aux pannes élevés. C'est pour toutes ces raisons que Cassandra est une base de données de ML et d'IA de tout premier plan. La preuve en est qu'elle est utilisée par Instagram, Netflix ou encore Reddit. Parmi ses principaux avantages, elle offre une mise à l'échelle horizontale linéaire, une BDD décentralisée avec réplication multicentres, ainsi qu'une tolérance aux pannes avec réplication automatique des données sur plusieurs nœuds.

Chorus.ai

Chorus.ai adopte une approche très spécialisée en visant les équipes en contact avec les clients et les équipes de vente. Elle fournit un assistant IA conçu pour améliorer les processus de prise de notes. À l'heure où les entreprises s'efforcent de rationaliser les interactions et de recueillir des informations sur les engagements des clients, l'assistant IA de Chorus.ai joue un rôle essentiel dans la capture d'informations importantes et la promotion d'une communication efficace.

Couchbase

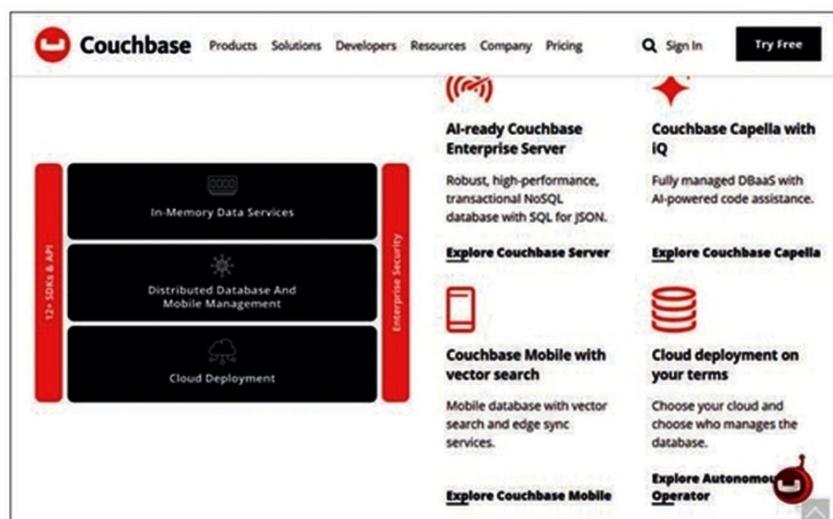
Couchbase (www.couchbase.com) est une BDD distribuée open source offrant une combinaison très puissante de haute disponibilité et de latences inférieures à la milliseconde. Au-delà de ses performances exceptionnelles, Couchbase intègre également des fonctionnalités Big Data et SQL qui la positionnent comme une solution totalement adaptée à des tâches complexes d'IA et de ML, notamment pour des applications nécessitant un accès aux données en temps réel et des capacités d'analyse évoluées. Ajoutons à cela que Couchbase facilite la création d'applications en fournissant des API de développement simples et puissantes dans différents langages de programmation, connecteurs et outils.

DynamoDB

Amazon DynamoDB (<https://aws.amazon.com/dynamodb/>) est une BDD multirégionale entièrement gérée, proposant une sécurité intégrée, une cache mémoire et une sauvegarde/restauration automatisée. Sa popularité est démontrée par le nombre important de grandes entreprises qui l'utilisent comme Airbnb, Samsung ou Toyota. Elle réalise les chiffrements « au repos » en vue de réduire la complexité nécessaire à la protection des données sensibles. Parmi ses principaux avantages, il faut citer son évolutivité et ses capacités en matière de réplication des données. Son stockage virtuel illimité permet de stocker des quantités illimitées de données.

Elasticsearch

Construit sur la base de code d'Apache Lucene, Elasticsearch est un moteur de recherche et d'analyse distribué open source full text doté d'une API REST qui facilite l'extraction d'informations en temps réel. Ses capacités s'avèrent inestimables dans les applications exigeant une récupération et une analyse rapides des données de journalisme et de surveillance. Grâce à ses prouesses d'interrogation en temps réel, Elasticsearch contribue de manière significative à l'amélioration des workflows d'IA et de ML. Il prend en charge tous



Couchbase intègre à la fois des fonctionnalités de type Big Data et SQL, ce qui en fait une BDD totalement adaptée à la gestion de tâches complexes d'IA et de ML

les types de données : numériques, textuelles, géospatiales, structurées et non structurées. Elasticsearch est intégré à Elastic Stack qui offre aussi différents outils open source pour l'enrichissement, l'ingestion de données, le stockage, la visualisation et l'analyse.

Google Cloud Bigtable

Google Cloud Bigtable (<https://cloud.google.com/bigtable>) est une base de données d'IA NoSQL distribuée, offrant une évolutivité robuste, une faible latence et la cohérence des données. Ces caractéristiques la rendent particulièrement apte à gérer les exigences d'accès aux données à haut débit. Néanmoins, et c'est l'aspect le plus négatif, la complexité de sa tarification nécessite une attention très particulière lors de sa mise en œuvre et est susceptible d'en décourager plus d'un.

LE MIT A CRÉÉ UNE BDD AUTO-ÉVOLUTIVE

Le laboratoire CSAIL (Computer Science and Artificial Intelligence Laboratory) du MIT a créé dès 2020 un système de stockage de données basé sur l'intelligence artificielle. Ce système de type « instance-optimized » est capable, grâce au ML (Machine Learning), d'apprendre au fil du temps à s'optimiser de manière autonome afin de traiter plus efficacement les requêtes des utilisateurs. L'approche du MIT consiste à créer des systèmes capables de s'optimiser et de se réorganiser partiellement de manière autonome en fonction des données stockées ou des workloads à traiter. Ils ont pour cela développé deux systèmes : Tsunami et Bao. Le premier, Tsunami, utilise le ML pour réorganiser automatiquement la disposition d'un ensemble de données en fonction du type de requête formulée. Cette approche permet d'exécuter les requêtes 10 fois plus rapidement que les systèmes traditionnels. Le second système, Bao, se focalise sur l'optimisation des requêtes. Il est capable d'apprendre au fil du temps en choisissant le meilleur plan de requête pour chaque situation. Bao a été pleinement intégré à PostgreSQL.

MongoDB

Le principal intérêt de MongoDB réside dans son approche flexible et orientée document en matière de gestion des données. Ce facteur associé à ses grandes capacités d'évolutivité en fait un choix intéressant pour la gestion de données non structurées. Les fonctionnalités de MongoDB séduiront sans nul doute les développeurs qui cherchent à gérer des structures de données complexes et à s'adapter à la nature dynamique de projets d'IA et de ML.

PostgreSQL

PostgreSQL, dans sa dernière version, est un système de base de données IA objet-relationnel open-source. Il a acquis sa réputation grâce à son respect indéfectible en matière d'intégrité des données et ses mécanismes d'indexation avancés. Il prend en charge divers types et structures de données. Définitivement ACID (atomicité, cohérence, isolation, durabilité), PostgreSQL est parfait pour gérer des charges de travail volumineux avec des données complexes tout en garantissant sécurité et fiabilité.

Redis

Redis (<https://redis.io/>) est une structure de données open source en mémoire plébiscitée par sa polyvalence et ses fonctionnalités très robustes. Elle prend en charge différents types de données, des simples chaînes de caractères aux structures de données les plus complexes comme les bitmaps, les ensembles triés, les index géospatiaux ou les hyperlogs. Redis englobe un large éventail de fonctionnalités allant de la prise en charge des transactions à la réplication des données, améliorant ainsi la durabilité et la disponibilité des données. Elle permet aussi de gérer des quantités massives de données, ce qui la rend totalement adaptée à l'IA, au Big Data et à l'IoT. Redis est principalement utilisée comme base de données clé-valeur distribuée en mémoire ou encore comme cache et courtier de messages. C'est à l'heure actuelle l'un des choix les plus populaires du

marché. Elle propose également des scripts Lua, une expulsion LRU, une réplication intégrée, des transactions ainsi que différents niveaux de persistance sur disque. Ses autres avantages qui la rendent incontournable sont son processus de basculement automatique et son module Redis-ML qui implémente plusieurs modèles d'apprentissage automatique en tant que types de données Redis intégrés.

MLDB

MLDB (Machine Learning DataBase) est un système open source destiné à la réalisation de tâches d'apprentissage automatique pour du Big Data. Il est employé pour collecter et stocker des données via la formation de modèles d'apprentissage automatique et pour déployer des points de terminaison de prédiction en temps réel. MLDB est l'un des ensembles de données les plus faciles à utiliser du fait qu'il fournit une implémentation complète de l'instruction SELECT du SQL (ce qui est très rare. Les aficionados du langage SQL vous le confirmeront). Il est donc capable de traiter des ensembles de données comme de simples tables. Cela facilite grandement le travail des Data Analyst connaissant bien les SGBDR (Système de Gestion de Base de Données Relationnelle) classiques.

MindsDB

MindsDB est une base de données virtuelle pour l'intelligence artificielle qui facilite la création d'applications centrées sur l'IA. Son objectif est de rendre l'IA accessible au plus grand nombre possibles d'entreprises en permettant à leurs développeurs d'intégrer des capacités d'IA dans leurs produits sans avoir à tout changer. MindsDB va, en gros, permettre d'ajouter des fonctionnalités de type IA/ML à des SGBD traditionnels, évitant ainsi aux développeurs de revoir complètement l'architecture de stockage de leurs applications. Jorge Torres, co-fondateur et PDG de MindsDB, a déclaré que « *bien que seulement 5% des développeurs soient experts en IA/ML, il est essentiel que chacun d'entre eux soit capable d'intégrer des modèles d'IA à leur infrastructure logicielle, problématique résolue par MindsDB* ». La société a annoncé avoir reçu récemment un investissement de NVentures, le bras de capital-risque de NVIDIA, portant le financement total en phase d'amorçage à 46,5 millions de dollars. Les investisseurs précédents comprennent Benchmark, Mayfield, Y Combinator, OpenOcean et Walden Catalyst. Nous parlerons plus en détail dans un prochain numéro de cette BDD « magique » qui permet d'adapter quasiment n'importe quelle BDD classique à l'IA. □



La gestion efficace par Amazon Aurora (<https://aws.amazon.com/rds/aurora/>) des charges de travail de données complexes en fait un choix privilégié pour les applications d'IA et de ML.

TT



ExpertBook B5 Flip

Puissance & mobilité pour votre entreprise



SÉCURITÉ
RENFORCÉE



FLEXIBILITÉ
ULTIME



PERFORMANCE
ÉLEVÉE



Windows 11

ASUS recommande Windows 11 Professionnel pour les entreprises

Vous ne me trouverez pas sur Amazon !



L'ouvrage de Laurent Mauduit, journaliste de Mediapart dont il a été un des fondateurs après avoir dirigé le service économique à Libération, est, comme il le dit lui-même, une « bouteille à la mer » face à la domination discrète, mais effective, des GAFAM sur la presse et l'édition grâce à leurs outils numériques de publicité. Il

n'y a pas que les milliardaires qui se disputent le marché de l'opinion. Ainsi, Amazon est devenu le premier distributeur de livres. Laurent Mauduit entend lancer l'alerte sur les dangers que les géants du numérique font peser sur la liberté de l'information dont la presse est l'instrument, et sur la liberté d'opinion dont le livre est

l'un des principaux vecteurs. Son appel est de ne pas pactiser avec les GAFAM et leurs instruments, et veut défendre une presse et une édition libre et indépendante. À quelques jours des élections européennes, il nous a paru salutaire de relayer cet appel et de faire passer le message contenu dans cette bouteille à la mer.

PRÉAMBULE

Aussi controversé qu'il soit, le géant Amazon est devenu en France le premier libraire et le premier distributeur de livres. Si tous les libraires indépendants connaissent les dangers que fait peser cette situation de domination sur l'acquis démocratique majeur que constitue le prix unique du livre et s'ils mènent de très longue date un combat solidaire pour contenir ses avancées, il n'en va pas de même des éditeurs dont beaucoup laissent faire. L'hypocrisie est même encore plus spectaculaire : de nombreux éditeurs de gauche, qui aiment publier des essais présentant les effets ravageurs du capitalisme financiarisé, n'ont aucune gêne à faire distribuer leurs livres par l'un des groupes les plus emblématiques de ce système prédateur.

Les éditeurs qui refusent ce double jeu sont rarissimes. Divergences est l'une des 80 maisons d'édition qui se sont fédérées autour d'Hobo Diffusion, une structure de diffusion de livres, DVD et revues en librairies, pour participer à l'appel de novembre 2020

« Nous ne vendrons plus nos livres sur Amazon ».

Comme cet essai, qui prolonge plusieurs de mes enquêtes publiées sur Mediapart, entend lancer l'alerte sur les dangers que les géants du numérique font peser sur la liberté de l'information dont la presse est l'instrument, et sur la liberté d'opinion dont le livre est l'un des principaux vecteurs, il m'a semblé logique qu'il soit publié par les éditions Divergences dont le fondateur, Johan Badour, a été l'un des rédacteurs de cet appel.

C'est de ce partenariat qu'est née l'idée du titre de ce livre : « *Vous ne me trouverez pas sur Amazon !* » Car les éditions neuves de cet ouvrage n'y seront effectivement pas distribuées : pour les trouver, il faudra se rendre dans une librairie. Mais ce titre est évidemment une bouteille à la mer. Nous nous prenons à espérer que le message finira par se propager sous des formes diverses, dans les univers du livre et de la presse : « *Nous ne pactiserons pas avec Google, pas plus qu'avec Facebook !* » C'est en tout cas l'ambition de cet essai : lancer l'alerte et défendre la presse et l'édition indépendantes.

INTRODUCTION

Aux géants du numérique, il est fait d'innombrables reproches. En vrac, celui de profiter trop souvent d'abus de position dominante et d'écraser la concurrence ; celui aussi de fuir l'impôt par des systèmes d'optimisation ou d'évasion sophistiqués ; celui également de ne pas respecter les règles sociales en vigueur en Europe continentale et particulièrement en France, et de promouvoir une ultra-flexibilité ; ou enfin de favoriser une captation générale des données de leurs utilisateurs et de les transformer eux-mêmes en marchandise, pour mettre la main sur une part toujours croissante du gigantesque marché publicitaire.

Mais à ces mêmes géants du numérique, on ne fait ordinairement pas grief de constituer des menaces directes contre les libertés fondamentales.

Et pourtant, c'est bel et bien ce constat que cet essai entend dresser : avec ces oligopoles anglosaxons, si puissants et si intrusifs, ce sont les libertés essentielles elles-mêmes qui

sont aujourd'hui en danger ; c'est en France la célèbre loi de 1881 les garantissant qui est menacée ; ce sont, en somme, des grandes conquêtes démocratiques, par exemple celles portées dans ce domaine en d'autres temps par Georges Clemenceau, qui sont en passe d'être balayées.

Certes, l'histoire est parfois injuste. Des grands hommes qui ont défendu au cours des derniers siècles la liberté de la presse et la liberté d'expression, on ne retient le plus souvent que quelques noms. Camille Desmoulins, qui voyait en la presse « la sentinelle » de la démocratie, est ainsi passé à la postérité pour avoir joué un rôle majeur dans la défense de ces principes dès le début de la Révolution française et contribué, avec l'appui de Robespierre, à la rédaction du célèbre article 11 de la Déclaration des droits de l'homme et du citoyen : « *La libre communication des pensées et des opinions est un des droits les plus précieux de l'homme : tout citoyen peut donc parler, écrire, imprimer librement, sauf à répondre de l'abus de cette liberté dans les cas déterminés par la loi.* »

Parmi ces grandes voix, il y a encore celle d'Alexis de Tocqueville qui, écrivant son ouvrage *De la démocratie en Amérique* dans les années 1835-1840, rappelle les principes sur lesquels la démocratie s'est construite au xviiie siècle, en France aussi bien qu'aux États-Unis. Et puis, il y a surtout la voix de Victor Hugo, la plus forte de toutes, la plus remarquable aussi, qui, avec vigueur, défend les valeurs universelles qui sont celles de la République, à l'occasion notamment de deux discours formidables, l'un prononcé le 11 septembre 1848, depuis la tribune de l'Assemblée constituante, et l'autre presque deux ans plus tard, le 9 juillet 1850, toujours depuis la tribune de l'Assemblée nationale. « *Le principe de la liberté de la presse n'est pas moins essentiel, n'est pas moins sacré que le principe du suffrage universel. Ce sont les deux côtés du même fait. Ces deux principes s'appellent et se complètent réciproquement. La liberté de la presse à côté du suffrage universel, c'est la pensée de tous éclairant le gouvernement de tous. Attenter à l'une, c'est attenter à l'autre* », proclame-t-il lors du premier discours.

Mais dans cette longue généalogie démocratique qui fait de la presse un fondement majeur de la démocratie, on omet le plus souvent de citer le nom de Georges Clemenceau. Sans doute y a-t-il une explication à cette omission. Dans la longue carrière politique qu'il connaît, on ne se souvient souvent que de son rôle détestable comme ministre de l'Intérieur, quand il réprime avec violence les grandes grèves des vignerons de l'année 1907 dans le Languedoc et le Roussillon ; ou son rôle de « Père la Victoire », quand il poursuit les défaitistes ou les pacifistes, ou mate les rébellions dans l'armée, à la fin de la Première Guerre mondiale. Et du même coup, on oublie ce qu'ont été sa jeunesse politique et ses combats démocratiques, du temps où, militant dans le camp des républicains, il se range du côté des « radicaux » et polémique contre les « opportunistes ». Et l'on oublie aussi le rôle majeur qu'il joue, contre les républicains modérés ou les monarchistes, dans le long et passionnant débat parlementaire qui conduit à la fameuse loi du 29 juillet 1881 sur la liberté de la presse et la liberté de parole.



Laurent Mauduit, journaliste de Mediapart dont il a été un des fondateurs après avoir dirigé le service économique à Libération.

Or, en ces temps obscurs dans lesquels la France est aujourd'hui en train de plonger du fait de la gouvernance autoritaire d'Emmanuel Macron, marqués par de très graves régressions démocratiques avec l'extrême droite en embuscade, il faut honorer les républicains à qui l'on doit la loi progressiste du 29 juillet 1881 — qui permet aussi de mieux comprendre les dangers que font peser ces géants du numérique sur ces libertés fondamentales, avec l'appui du même Emmanuel Macron.

On oublie trop souvent que les révolutionnaires de 1789, en proclamant la Déclaration des droits de l'homme et du citoyen, ne posent qu'un principe fondateur : la liberté de la presse et la liberté de parole sont les premiers ressorts d'une démocratie — ce que les pères fondateurs de la démocratie américaine établissent eux aussi, à la même époque, avec le célèbre Premier amendement. Mais par la suite, à partir de 1793 et jusqu'aux débuts de la IIIe République, la liberté de la presse et la liberté d'expression connaissent une longue période de glaciation, avec quelques très courtes périodes d'éclaircies démocratiques, en 1830 ou en 1848. Il faut donc attendre 1881, presque un siècle plus tard, pour que l'article 11 de la Déclaration des droits de l'homme trouve enfin une traduction législative. Cette loi de 1881, il est d'autant plus utile d'en souligner l'importance qu'elle fait l'objet, depuis quelques années, singulièrement depuis le début du premier quinquennat d'Emmanuel Macron, d'innombrables attaques et remises en cause. Et au nombre de ces attaques, il y a celles que sont en train de mener les GAFAM — acronyme de Google, Amazon, Facebook, Apple, Microsoft, c'est-à-dire les cinq conglomérats américains qui dominent le marché mondial du numérique. Cette offensive risque d'ébranler

encore un peu plus cette loi de 1881, qui est la principale conquête démocratique garantissant le droit de savoir des citoyens. En se remémorant le combat des républicains de 1881, on devine ce que la France pourrait perdre si cette conquête démocratique était remise en cause.

Si Georges Clemenceau s'engage de la sorte dans le combat pour la liberté de la presse et la liberté de parole en 1881, c'est qu'une décennie plus tôt, de 1865 à 1869, il a séjourné aux États-Unis et y a été le correspondant du journal *Le Temps*. Durant ces années, au cours desquelles il a découvert les vertus libérales de la démocratie américaine et du rôle de la presse, il s'est forgé de fortes convictions démocratiques, qui transparaissent quand il est question de faire en 1881 une loi sur la liberté de la presse et la liberté de parole. La détestation de l'empire et la sympathie pour la démocratie américaine le conduisent à s'engager fortement.

Alors que l'examen du projet de loi par la Chambre des députés a commencé depuis plusieurs jours, c'est lors de la séance du 5 février 1881 que Georges Clemenceau fait un premier long discours afin de défendre ses positions. Pour commencer, il rappelle le prix que son père, Benjamin Clemenceau, envoyé en prison à plusieurs reprises sous le Second Empire, a payé pour ses engagements républicains : « *J'ai été élevé dans l'amour de la République. J'ai vu mon père frappé au Deux-Décembre. Plus tard, je l'ai vu partir pour l'Afrique, enchaîné comme un malfaiteur.* » Et aussitôt, il en tire argument pour une ode à la liberté : « *Il n'y a pas un acte de ma politique où je n'aie vivement cherché à servir la cause républicaine. Et je viens défendre ici le principe de la République, la liberté [...] contre quelques-uns de ceux qui ont aujourd'hui la prétention de défendre la République mieux que nous ne saurions le faire et qui pourtant servaient Bonaparte quand les républicains étaient proscrits [...] La République vit de liberté ; elle pourrait mourir de répression, comme tous les gouvernements qui l'ont précédée et qui ont compté sur le système répressif pour les protéger. [...] On veut vous faire décréter que l'outrage à la République est punissable. Soit. Et puis qu'arrivera-t-il ? Il y aura des magistrats qui siégeront dans le prétoire et qui auront pour mission de dire où finit la discussion et où commence l'outrage.* »

Et, prolongeant son remarquable discours, Clemenceau ajoute : « *Qu'est-ce donc, en effet, que la République, sinon un gouvernement d'opinion, c'est-à-dire un gouvernement fondé sur le principe du respect de la volonté nationale et reposant par conséquent sur le principe de la liberté complète de discussion [...] Ce qu'on vous demande de faire est possible dans un gouvernement monarchique, où le principe, c'est le roi ou l'empereur. On peut à la rigueur déterminer ce que c'est que l'outrage à la personne du roi ou de l'empereur, qui se confond avec le principe. Mais aujourd'hui, où est le roi, où est l'empereur ? [...] Le souverain, c'est le suffrage universel, c'est la nation, gouvernant dans sa liberté et trouvant dans la forme républicaine la seule garantie de cette liberté.* »

Le propos est bien sûr marqué par le souvenir de ce qu'a été le régime autoritaire de Napoléon le Petit, mais il garde avec le recul une résonance beaucoup plus profonde sur ce qu'est la liberté d'expression.

« *Vous n'y songez pas, ajoute Clemenceau, vous avez vaincu vos ennemis, les ennemis de la République, et cependant vous aviez contre vous les puissances de ce pays, l'administration, la magistrature, l'armée elle-même spécialement organisée contre le pays, l'argent lui-même, l'oligarchie financière ; vous aviez contre vous les puissances sociales, armées de toutes les lois de répression dirigées par toutes les monarchies contre la liberté. Et cependant, vous avez vaincu ! Et ces lois de répression ont été impuissantes, et votre victoire a consacré à jamais le triomphe de la liberté et l'impuissance du système répressif. Et c'est quand votre victoire est assurée, quand vous êtes les plus forts, quand vous avez une majorité immense non seulement dans le Parlement, mais dans le pays, quand votre pouvoir prouve l'inefficacité de ces mêmes lois qu'on vous propose d'édicter aujourd'hui, quand vous avez pris sur ces bancs la place de ceux qui défendaient ces lois.* » □

VOUS NE ME TROUVEREZ PAS SUR AMAZON !

Les milliardaires ne sont pas les seuls à avoir conduit une opération de predators sur la presse et l'édition. Amazon est devenu le premier distributeur de livres. Google et Facebook sont devenus des acteurs majeurs de la presse via leurs réseaux publicitaires. S'il y a urgence à faire barrage à Bolloré et consorts pour défendre le droit de savoir des citoyens, il faut aussi combattre ces géants du numérique qui veulent remettre en cause le prix unique du livre et asservir la presse. À travers une enquête sur le monde de l'édition et les médias, Laurent Mauduit nous montre les dangers que les géants du numérique font peser sur la liberté d'information et la liberté d'opinion.

LAURENT MAUDUIT

EAN : 9791097088675 — 150 pages
ÉDITIONS DIVERGENCES (23/02/2024)

éditions divergences



1^{er} Club Français de décideurs informatiques & télécoms

Un réseau privé et indépendant
où siègent 13 DSi et 1 RSSI
1500 Membres

Le Club accompagne les DSi à faire les bons choix technologiques
en adéquation avec leurs projets.



FONDATEURS



Véronique Daval Présidente Julien Daval Vice-Président

LES MEMBRES DU BUREAU ET AMBASSADEURS DU CLUB



Armand ASSOULINE
CIO & National
Documentation
Manager - MSC



Laurent BAYOL
DSi
LA COMPAGNIE



Nawal BENASSI
CIO & Digital Officer
ESRI FRANCE



Gilles BERTHELOT
Directeur Sécurité
Numériques
GROUPE SNCF



Christophe BOUTONNET
Directeur Adjoint
du Numériques
Ministère écologie,
énergie,
territoires et mer



Benoit DECOCK
Business transformation
numérique leader
AGFA



Christian DOGUET
CIO
CHAINE THÉRALE
DU SOLEIL



Alain GUEDE
CIO
GROUPE SARETEC



Christophe GUILLARME
RSSI
LAGARDÈRE TRAVEL
RETAIL



Philippe LAGRANGE
Directeur recherche
et prospective
MUTUELLE GÉNÉRALE
DE LA POLICE



Stéphane MALGRAND
DSi
LABORATOIRE NATIONAL
MÉTROLOGIE
ET ESSAIS



Sandrine RACOUCHOT
DSi
INTER MUTUELLES
HABITAT



Lionel ROBIN
DSi
THE SET hOTELS



Claude YAMEOGO
ARCHITECT SI
ALSTOM



COORDINATEUR



TRIEU Huynh-Thien

CLUB DECISION DSi 33, Rue Galilée 75116 Paris • Tél +33 1 53 45 28 65
Contact : Véronique DAVAL - Présidente • veronique.daval@decisiondsi.com

www.clubdecisiondsi.fr



Shoptalk à Las Vegas

L'IA prend place dans le retail

Grand rendez-vous du retail, le salon Shoptalk s'est tenu à Las Vegas du 17 au 20 mars.

Pour cette édition, de nombreuses entreprises ont présenté leurs technologies pour le retail. Mais c'est clairement l'intelligence artificielle qui était dans toutes les conversations.

Organisé aux États-Unis et en Europe, Shoptalk est la grande-messe du retail. C'est à Las Vegas que s'est tenue la session du printemps (17-20 mars) avec plus de 5 000 visiteurs. Des acteurs comme SAP, Salesforce ou Adobe ont montré comment ils intervenaient dans le retail, mais aussi quelles solutions ils proposaient aux entreprises. SAP a ainsi annoncé le lancement d'une nouvelle solution de paiement composable avec SAP Commerce Cloud avec des fournisseurs de services (Stripe, Adyen, WorldPay et Airwallex). Cette architecture spécifique permet notamment aux utilisateurs de choisir des partenaires de paiement adaptés à leurs besoins et à leurs marchés internationaux. L'objectif étant aussi de réduire la complexité en éliminant le besoin d'intégrer et de déployer des extensions. Cette nouveauté est disponible en version bêta avec un lancement officiel prévu au second semestre. On peut aussi relever la présentation d'Adobe Commerce (basé sur Adobe Experience Cloud) pour assurer la gestion jusqu'au paiement final par le consommateur.

Dans un autre registre, la société américaine Zonos a présenté ses services, donnant la possibilité aux entreprises (via des API) de calculer automatiquement les frais de port, les tarifs douaniers et les taxes dans le monde entier. Parmi les autres exposants, Wiliot se focalise sur les IoT en développant des étiquettes intelligentes proches de la technologie RFID. Mais l'entreprise va plus loin. En se connectant à un boîtier Bluetooth, ces étiquettes dotées de puces (fabriquées par Avery Dennison et Identin) livrent de nombreuses informations sur les produits pour savoir s'ils sont périmés ou bien pour connaître l'état des stocks via une plateforme SaaS. « Il faut compter quelques cents pour une étiquette qui pourra être mise sur un produit ou sur des cartons », explique Zach Lowman, membre de l'équipe marketing de Wiliot.

L'intelligence artificielle au soutien du retail

Autre enseignement : l'intelligence artificielle voit ses applications se multiplier dans le commerce. Crée il y a moins d'un an, la société WorkMagic a développé une plateforme en ligne et de nombreux outils utilisant l'IA et le machine learning (ML). L'outil Marketing Copilot est un assistant, basé sur plusieurs agents IA, aide à gérer la génération de contenu, le développement de campagnes, l'analyse d'attribution et les connaissances générales sur l'e-commerce. Connue pour



Durant quatre jours, les professionnels du retail se sont retrouvés à Las Vegas pour assister à Shoptalk. Ils ont pu rencontrer des entreprises spécialisées dans le développement de solutions pour le back office, la gestion des stocks et de la clientèle, avec en filigrane l'intensification de l'utilisation de l'intelligence artificielle.

ses interfaces de messaging, Twilio a montré les capacités de sa technologie CustomerAI qui associe la puissance des grands modèles de langage (LLM) aux données clients en temps réel qui transmettent sur sa plateforme. Le but est de permettre aux entreprises de mieux comprendre et exploiter le potentiel des clients avec une valeur ajoutée grâce aux capacités prédictives de l'IA. « L'intelligence artificielle représente le futur. Nous allons fournir encore plus d'efforts sur ce secteur cette année », souligne Mark Morin, ingénieur en chef de Twilio. De son côté, la start-up israélo-américaine Tymely utilise également l'IA pour le service client. « Notre agent IA n'est pas un chatbot. Il s'agit d'une technologie qui permet d'automatiser et d'améliorer l'expérience du service client, en atteignant 60 à 80 % d'automatisation tout en augmentant de manière significative le taux de satisfaction client », souligne Josh Behar, vice-président comptes stratégiques et développement. Persado, une entreprise qui a développé sa propre intelligence artificielle (Motivation AI), propose notamment des solutions pour le marketing numérique. La société se distingue par une plateforme qui se base sur des interactions réelles et des données de transaction provenant de 1,2 milliard de consommateurs. Selon l'entreprise, qui compte Carrefour parmi ses clients, ce mix améliore en moyenne de 40 % les performances. Il y a quelques jours, Persado a gagné le prix Artificial Intelligence Excellence Awards Programm dans la catégorie IA générative.

Bref, cette édition de Shoptalk a placé l'intelligence artificielle au cœur de toutes les réflexions et de tous les développements. L'IA se retrouve partout, que ce soit pour aider les entreprises du retail ou pour apporter une nouvelle expérience d'achat aux consommateurs. Et ce n'est que le début. □

Michel Chotard

Confiance.ai

Doper l'ingénierie avec de l'IA

Avec le consortium Confiance.ai, les industriels avancent sur la mise en œuvre et l'industrialisation de composants d'IA dans les systèmes critiques.

La robustesse, l'explicabilité et la transparence de ces technologies font l'objet d'une attention particulière, notamment pour des raisons sociétales.

Des défis de taille sont encore à relever.

ancé en 2019 pour 4 années, le programme Confiance.ai cofinancé par l'État et ses partenaires industriels va bientôt se clore. Ses responsables ont organisé un évènement au MINEFI, le 7 mars dernier, pour présenter les résultats et les défis en cours et à venir. Ce programme avait pour but de faciliter l'utilisation de composants d'IA dans des systèmes critiques. Ce qui suppose que les industriels puissent les utiliser avec des garanties en particulier de sécurité, de robustesse et d'explicabilité. « *Sans IA responsable et sans confiance, il n'y a pas de projets d'IA possibles pour des applications critiques* », ont souligné les intervenants Bernhard Quendt, CTO de Thalès, Mohammed Sijelmassi, CTO de Sopra Steria, et Baladji Soussilane, VP Digital et IT Group d'Air Liquide lors de la première table ronde. Dans cette logique, l'objectif du programme était donc de pouvoir formaliser des méthodes et d'identifier des outils chargés de certifier des composants d'IA dits « de confiance ». Il s'agissait également d'outiller les projets pour faciliter l'industrialisation et la maintenance de ces composants. Enfin, l'objectif restait de proposer des solutions aussi génériques que possible aptes à s'adapter à tous les contextes industriels.

Les équipes de Confiance.ai, issues de SystemX et de partenaires industriels et académiques, étaient constituées pour chaque cas d'usage. Elles ont commencé par lister les processus d'ingénierie, et par les « revisiter » pour éventuellement pouvoir intégrer des composants d'IA de la même manière que tout autre composant. Les algorithmes retenus étaient de l'apprentissage machine dans quasiment tous les cas. Cette approche a nécessité de faire travailler ensemble et d'adopter une terminologie commune par de nombreux métiers : data scientist, responsable de la sécurité, de la fiabilité... Le but était de formaliser les propriétés de confiance attendues, robustesse par exemple, de définir la méthode et d'identifier les outils capables de les garantir. Une démarche qui est passée par l'identification des algorithmes open source existants ou d'outils propriétaires, souvent développés par les start-ups partenaires du programme. Le projet s'est concrétisé entre autres par la définition d'un état de l'art des outils et de benchmark. Exemple, un des composants dénommé Kaa prend la forme d'une plateforme embarquant des plugins d'explicabilité comme Captum pour tester les modèles. Le programme n'avait pas vocation à développer des algorithmes ou des composants. « *Nous avons adopté une approche de bout en bout, a souligné Juliette Mattioli, présidente du comité de pilotage. Depuis la conception de composants d'IA dans un système critique à sa maintenance en passant par la validation de jeux de données, le monitoring... ainsi que « l'embarquabilité ».* Ce dernier point est souvent un facteur déterminant dans certains cas d'usage, défense et aéronautique notamment, qui imposent des contraintes en termes de ressources.

Une dizaine d'industriels ont participé au projet en apportant leurs propres cas d'usage. Ces derniers ont servi de banc d'essai. Pour assurer le caractère générique des méthodes et des outils, ceux-ci ont été testés sur un autre cas d'usage que celui à partir duquel ils ont été conçus. Ces cas d'usage portaient majoritairement sur

© GilLefauconnier



la reconnaissance d'images et sur l'analyse de séries temporelles. Parmi les cas présentés, Air Liquide considère comme une donnée critique l'état des stocks de bouteilles d'oxygène. Préalablement au projet, une application de reconnaissance d'images basée sur l'IA et embarquée dans des caméras sur des portiques comptabilisait ces bouteilles sur les camions quand ces derniers rentraient dans les usines. Les résultats restaient dépendants des conditions météorologiques et de la luminosité. Le but était d'améliorer ce comptage. Les travaux se sont attachés à améliorer la qualité des données, à savoir pré-entraîner les algorithmes avec des données pour éliminer les gouttes de pluie, les flocons de neige ou encore, transformer des photos de nuit dans leur équivalent de jour. Avec cette méthode, la précision de comptage est passée à 98 % et le nombre d'erreurs a baissé de 50 %. Les problématiques plus spécifiquement analysées dans ce contexte ont été la gestion du cycle de vie de la donnée, la robustesse et la supervision. De son côté, Renault a travaillé sur la vérification de la qualité des soudures critiques sur les véhicules, une tâche jusque-là dévolue à un collaborateur. Une application existait déjà, mais ses résultats, une précision de 97 %, était considérée par le constructeur comme insuffisante pour un déploiement opérationnel. Ce projet a ciblé la robustesse, l'explicabilité et les fonctions de monitoring. Toujours dans le domaine de la reconnaissance d'images, un composant impliquant Airbus a pour but une aide à l'atterrissement, et Valéo, des aides à la conduite. Dans ce dernier cas, l'entraînement a été effectué avec des images réelles et d'autres générées synthétiquement. Data Scientist sur le projet, Amélie Bosca détaille : « *il reste des erreurs. Exemple, sur une photo de nuit, l'algorithme a transformé un poteau en feu de signalisation* ». Autre défi, les données utilisées pour les étapes d'entraînement des réseaux ne sont plus forcément en phase avec la réalité. En d'autres termes, le domaine opérationnel a évolué. Exemple, pendant la période du Covid, l'IA a vu apparaître des piétons masqués. Pour garantir des

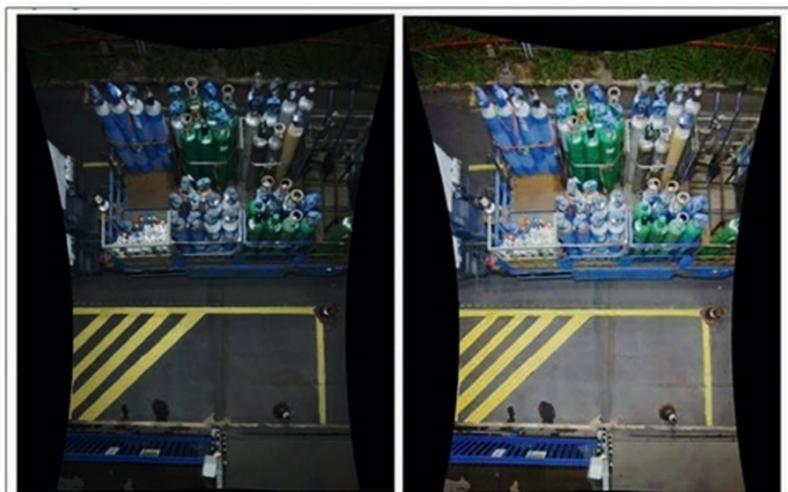


Photo transformée par l'IA pour améliorer le comptage.

résultats acceptables, des contrôles basés cette fois-ci sur des systèmes classiques de règles sont chargés de sécuriser les composants.

Des résultats en ligne

Considérés comme un « bien commun » numérique destiné avant tout à booster l'industrie française et européenne, les résultats du projet, méthodes et logiciels, sont disponibles sur le site www.confiance.ai. L'IA générative n'est pas incluse dans le catalogue. « *Les LLM ne cochent aucune case de l'IA de confiance et restent attaquables* », a souligné Fabien Mangeant, directeur scientifique informatique et data science chez Air Liquide et président du comité de direction de Confiance.ai. Pour autant, plusieurs projets, Air Liquide compris, ont fait appel à cette technologie. Elle n'a été utilisée que pour enrichir les jeux de données utilisées pour l'apprentissage, par exemple pour générer des images avec des gouttes de pluie, de la neige... Le site propose également les résultats scientifiques, à savoir les articles publiés, et technologiques. Restriction prévisible, certains composants ne sont pas accessibles en ligne, certains industriels en gardant pour eux une partie. Au total, un catalogue de 126 composants est proposé, les regroupant en ensembles fonctionnels. On y trouve des librairies python pour la robustesse, le monitoring, la gestion du cycle de vie de la donnée, l'explicabilité, la quantification de l'incertitude, ou encore les données synthétiques.

Les méthodes (sur <https://bok.confiance.ai/>) proposées incluent la formalisation des propriétés de confiance, robustesse... et les workflows chargés de les valider. Si le projet se termine en septembre, ses responsables espèrent un financement pour une V2. Ils ont noué plusieurs partenariats à l'international, avec l'organisation allemande VDE et avec un programme similaire québécois. Le projet est présenté comme une initiative visant à rassembler des acteurs du monde entier autour d'une feuille de route commune pour l'IA. □

LE DERNIER POURCENT TRÈS « CRITIQUE »

Outiler les industriels pour pouvoir utiliser des composants de machine learning dans des systèmes critiques, dans des avions, sur des missiles... avec des garanties, et de la même façon que d'autres composants, représente bien sûr une avancée. Les cas d'usage présentés, Air Liquide ou Renault, soulignent les avancées. Reste qu'une précision de 98 %, ou même 99 % est encore très loin des exigences habituelles de sécurité dans des domaines comme l'aéronautique (une erreur sur un milliard). Les défis technologiques sont encore de taille avant d'envisager une banalisation notamment pour du temps réel. Sauf, bien sûr, si l'on accepte un risque d'erreur plus élevé comme c'est certainement déjà le cas dans d'autres pays pour l'armement par exemple.

Pbr

Réindustrialisation

DeepTech 5 ans après

La BPI vient de faire le bilan du plan DeepTech lancé il y a 5 ans par le gouvernement dans le but de faire de notre pays un acteur majeur de l'innovation de rupture à l'échelle internationale.

Doté à l'origine de 3 milliards d'euro de financement, le plan DeepTech porte trois ambitions. La première est de stimuler la création de start-up sur des technologies de rupture avec un objectif de 500 jeunes pousses créées par an. Il vise aussi à accompagner la croissance et construire les leaders industriels de demain par une mobilisation massive de moyens, avec pour objectif l'émergence de 10 licornes d'ici 2025 et 50 sites industriels par an à l'horizon 2030. De plus, le plan a pour but de dynamiser les écosystèmes d'innovation par filières et dans les territoires.

Un bilan plutôt positif

Si tous les objectifs ne sont pas encore remplis, le plan a permis la création de plus de 1300 start-ups. Depuis la mise en place du Plan DeepTech en 2019, le nombre de start-up créées chaque année a doublé passant de 168 en 2018 à 340 en 2023. Au total, la France compte 2170 start-ups sur le territoire. La moitié d'entre elles sont situées dans les deux pôles majeurs que sont l'Ile-de-France et la Région Auvergne-Rhône-Alpes. Les startups deeptech exploitent 143 sites industriels dont 26 créés en 2023 (+73% par rapport à 2022), fortement liés aux thématiques d'innovation au niveau local. Par exemple, ROSI Solar a installé un site industriel en Isère pour recycler le silicium des panneaux solaires, Quandela a lancé un site sur le plateau de Saclay pour fabriquer des ordinateurs quantiques, et la startup Lhyfe a contribué à la naissance de deux nouveaux sites de production d'hydrogène vert dans le Morbihan et en Haute-Garonne. Le plus encourageant reste que cela a permis de créer des acteurs de stature internationale dans plusieurs secteurs stratégiques comme les batteries, la santé ou dans des secteurs d'avenir comme le quantique ou l'intelligence artificielle. Cela s'accompagne d'une forte mobilisation européenne qui enregistre une croissance continue des montants levés depuis 8 ans. En 2023, la résilience de l'écosystème français et son attractivité auront permis de se rapprocher du Royaume-Uni.

Une structuration de l'écosystème de l'innovation

En 2023, le développement de la deeptech issue du monde académique a franchi une étape importante, grâce à la création de 29 Pôles Universitaires d'Innovation (PUI),



Une vue de la giga factory de batteries électriques de Verkor qui est soutenue par le plan DeepTech.

qui disposent au total de 165 M€ de France 2030 pour renforcer la coopération entre l'université et les acteurs de l'innovation au niveau local, ainsi qu'à la Bourse French Tech LAB (BFT Lab), qui dispose de 20 M€ au total, un outil de financement mis en place par Bpifrance pour soutenir les projets les plus prometteurs directement dans les laboratoires et conçu comme un accompagnement à la maturation des projets deeptech, axé sur les aspects business de ceux-ci (études de marché(s), élaboration du modèle économique, compréhension des usages et besoins, ...).

Une forte participation du capital-risque

Dans un contexte global de difficulté des startups à se financer, les investissements dans la deeptech ont continué leur croissance en 2023, en hausse de 38% par rapport à 2022, portant les montants totaux levés à 4,1 Mds€ (soit 4 fois plus qu'il y a 5 ans). Cette dynamique est notamment portée par les giga-levées de Verkor (850 M€) et Mistral (498 M€). Les levées de fonds deeptech ont représenté 50% du marché du capital risque en 2023 contre 23% en 2018. Par la réorientation des investissements sur la deeptech, BPI a joué un rôle important en particulier par son fonds de fonds avec 397 M€ fléchés sur des fonds deeptech en 2023 (1,6 Md€ au total depuis 2019).

Au bilan, un écosystème multi-acteurs se structure, permettant de faire circuler les expertises et les talents. À ce titre, le lancement de France DeepTech, en fin d'année 2023 a permis de regrouper les principaux investisseurs deeptech et acteurs de la recherche, autour d'une volonté de collaborer pour structurer une communauté engagée autour des entrepreneurs deeptech. De la même manière, le lancement du fonds Sorbonne Ventures a montré la capacité des investisseurs et acteurs académiques à grouper leurs compétences. □

B.G

Leader dans la détection des cybermenaces, Gatewatcher protège depuis 2015 les réseaux critiques des grandes entreprises et des institutions publiques à travers le monde.

Nos solutions de Network Detection and Response (NDR) et de Cyber Threat Intelligence (CTI) détectent les intrusions et répondent rapidement à toutes les techniques d'attaque. Grâce à l'association de l'IA à des techniques d'analyse dynamiques, Gatewatcher offre une vision à 360° et en temps réel des cybermenaces sur l'ensemble du réseau, dans le cloud et on premise.

Sucess story

2015-2021

Création de Gatewatcher et établissement sur le marché national

2022-2024

- Expansion à l'international
- Levée de fonds, Série A
- Alliance européenne avec Open XDR
- Normé "Fournisseur Représentatif" dans le Market Guide 2022 for Network Détection & Response par Gartner®
- ouverture de bureaux au Royaume-Uni, Benelux, Singapour, EMEA
- Lauréat du programme French Tech 2030

Expertise

#NDR

#CTI

#TAP

#IOT

4,7/5



Peer Insights

In Network Detection and Response rating

16 ratings as of date 01/2024*

Produits



Gatewatcher NDR est une plateforme ouverte de détection et de réponse réseau offrant une cartographie à 360° des assets présents sur le SI et l'analyse contextuelle des cybermenaces pour une détection instantanée et une visibilité augmentée.



Gatewatcher CTI est un service de Threat Intelligence accessible par Feed, TIP ou plugin navigateur qui permet en un temps très court la détection des menaces internes et externes en fournissant des indices de compromissions enrichis et contextualisés à votre activité.

Use cases



Développer et améliorer mes capacités de détection et de protection



Identifier proactivement les menaces cyber



Adapter la détection à mes environnements spécifiques



Maîtriser et contrôler mes usages et mes données



Connaître les usages numériques de mon organisation



Bâtir mon écosystème cyber sur la base d'un NDR



Protéger mes environnements cloud



Grâce à un traitement machine learning et à une analyse contextuelle approfondie des comportements des utilisateurs, nos solutions permettent la détection des cybermenaces directement sur le réseau et simplifient le travail d'investigation de vos analystes.

Jacques de La Rivière - CEO



Campus Cyber - Tour Eria,
5 rue Bellini 92800 Puteaux, FRANCE

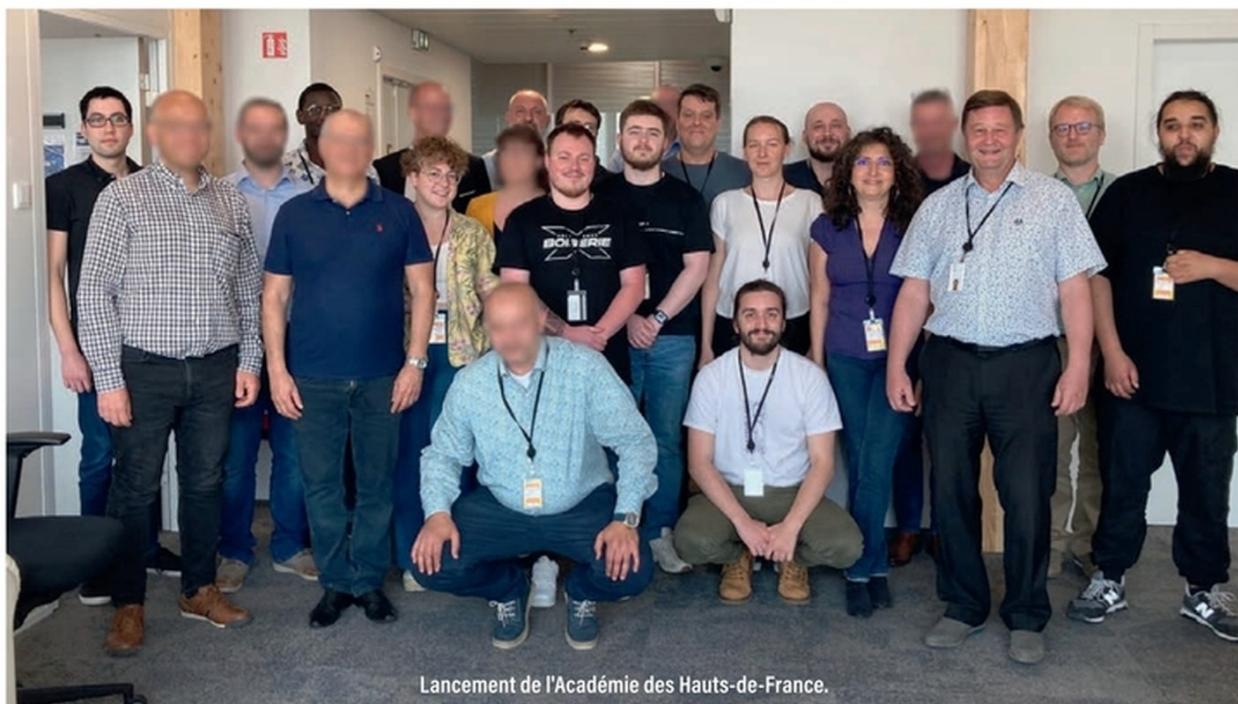
contact@gatewatcher.com
gatewatcher.com



Administration système

IBM déploie ses ailes pour former toujours plus d'administrateurs mainframe

Pour répondre à une demande constante en matière de ressources mainframe qui peine à être satisfaite, IBM vient de déployer plusieurs programmes, aussi bien au niveau mondial qu'hexagonal. S'il semble que les développeurs Cobol soient de plus en plus nombreux, c'est du côté des administrateurs z/OS que la pénurie est la plus forte.



Si le mainframe est loin d'être mort, il manque toutefois cruellement des personnes qualifiées pour le maintenir en vie. C'est en tout cas le constat, peu surprenant, d'une étude commandée par IBM et Broadcom sur le sujet. Celle-ci met ainsi en évidence une demande accrue des entreprises pour des profils orientés mainframe qui, même si les choses tendent à s'améliorer quant à la disponibilité des talents, n'est toujours pas satisfaite. En 2023, 32 % des employeurs dans le domaine du mainframe ayant répondu à cette enquête déclaraient avoir recruté entre 11 et 20 postes liés au mainframe et 35 %, plus de 20. En outre, 91 % de ces mêmes employeurs ont assuré qu'ils prévoyaient d'embaucher de nouveaux talents dans le domaine du mainframe au cours des 1 à 2 prochaines années, afin « d'investir dans une équipe informatique spécialisée dans les mainframes et la développer ».

« *Quand on parle de compétences mainframe, il faut aujourd'hui faire la différence entre deux profils très différents. D'un côté, nous avons les développeurs Cobol, et de l'autre côté, les administrateurs qui maîtrisent z/OS. Et c'est surtout ces derniers qui font défaut aujourd'hui* », résume Guillaume Wazer, Principal Leader Technique sur la plateforme Mainframe pour IBM France. Ainsi, la formation de développeurs Cobol a largement repris en France après avoir connu une certaine baisse, il y a une dizaine d'années. Les écoles d'ingénieurs et de développement proposent des modules Cobol qui sont aujourd'hui suivis par de nombreux étudiants. « *À cela, il faut ajouter que nous améliorons constamment l'expérience développeurs pour leur permettre d'utiliser d'autres langages* », ajoute Guillaume Wazer qui rappelle également que l'assistant Watson X permet maintenant de s'appuyer sur l'intelligence artificielle pour coder plus simplement ou faire du retro engineering d'autres langages.

Un focus sur les administrateurs

Non clairement, la principale préoccupation des entreprises qui s'appuient sur des mainframes n'est pas tant du côté du développement que de l'administration. « *Malgré les efforts que nous fournissons pour rendre la plateforme plus simple à gérer, nous manquons encore aujourd'hui cruellement d'administrateur z/OS* », explique Guillaume Wazer. Pour remédier à cette pénurie, IBM n'a cessé de déployer des initiatives. La dernière en date consiste en la mise en place du Mainframe Skills Council. Annoncée lors de l'événement Share qui s'est tenu à Orlando en mars dernier, cette initiative a pour objectif de réunir des organisations qui partagent leurs expériences et mettent en œuvre des solutions de manière collaborative afin de constituer une main-d'œuvre solide pour la plateforme mainframe. Le conseil comprendra ainsi des groupes de travail axés sur la connaissance des carrières, les cadres de compétences, les parcours de formation et le développement professionnel.

Le Mainframe Skills Council s'accompagne du lancement par IBM du Z Mainframe Skills Depot, une plateforme permettant d'accéder à la formation aux compétences mainframe. Celle-ci regroupe aujourd'hui plus de 700 heures de formations accessibles pour tous. Les apprenants peuvent



Guillaume Wazer, Principal Leader Technique sur la plateforme Mainframe pour IBM France.

ainsi choisir parmi des parcours spécialisés adaptés à des rôles tels qu'administrateur système, développeur d'applications ou architecte de modernisation. La plateforme donne par ailleurs accès à plus d'un millier d'heures de formation pratique, gratuite et à leur rythme, afin d'obtenir des badges numériques reconnus par le secteur.

Des programmes déjà solides en France

« *Ces nouveaux programmes vont venir enrichir les ressources que nous avons déjà déployées sur le marché français* », note Guillaume Wazer. Dans l'Hexagone, la firme d'Armonk a déjà mis en place plusieurs initiatives avec notamment la Z Académie de l'Intelligence Artificielle et Grand Système dont la première promotion, qui s'est tenue dans les Hauts-de-France, vient d'être diplômée. Cette formation portée conjointement par France Travail, IBM et le partenaire NormQuest a permis la certification et l'embauche de 15 premiers administrateurs mainframe suite à un an d'alternance entre salles de classe et entreprises. « *Les retours des entreprises qui ont participé ont été excellents. Toutes ont embauché les participants étant allés jusqu'au bout de la formation* », raconte Guillaume Wazer.

La deuxième promotion a débuté pour sa part en Ile-de-France depuis le mois de février avec une fois de plus, une quinzaine de participants, et d'autres devraient suivre à l'avenir. Ces formations s'adressent principalement à des profils en reconversion avec, de préférence un bagage scientifique et technique. Sur des cursus plus académiques, IBM a également noué des partenariats avec plusieurs universités et écoles françaises, notamment l'ENSISA Mulhouse, l'EPITA Paris et Telecom Nancy.

Une simplification par la technologie ?

« *Si la formation est une chose pour réussir à pourvoir suffisamment d'administrateurs, nous devons également rapprocher z/OS des standards de l'industrie afin d'en faciliter la prise en main* », ajoute Guillaume Wazer. IBM a ainsi travaillé à apporter une couche d'abstraction à même de simplifier l'usage des mainframes en les rendant compatibles avec des solutions comme Ansible, OpenShift ou plus simplement du Linux en fonction des types de déploiements visés ». Aujourd'hui, on tente de plus en plus de faire converger les mainframes vers les environnements cloud natif. Toutefois, nous aurons beau nous rapprocher des standards du marché, il y aura toujours des spécificités de z/OS qui nécessiteront des compétences particulières », détaille le responsable.

Quant à accroître l'usage de l'IA pour simplifier la gestion des mainframes, la dernière version de z/OS en date, la 3.1, intègre plusieurs fonctionnalités, notamment en matière de workload manager. « *Cela reste toutefois qu'un assistant. Nous ne remplacerons jamais les administrateurs systèmes par de l'IA* », précise Guillaume Wazer.

O. Ba

Tenacy

ALL YOUR CYBER IN ONE PLACE

tenacy.io

IA et cybersécurité

Une affaire de gentils et de méchants

Sommaire

IA et cybersécurité : une affaire de gentils et de méchants	P74
Checkpoint mise sur l'effet plateforme	P78
Sophos : Comment la GenAI va épauler les escrocs	P82
Cyber Resilience Act : l'UE muscle ses règles en matière de cyber	P85
Le chiffrement entièrement homomorphe de plus en plus concret	P87
Rencontre avec Guillaume Lopes, Senior Penetration Tester et CTO chez RandoriSec	P90

Difficile de passer à côté de l'intelligence artificielle ces derniers temps. Difficile aussi, de ne pas lui consacrer un dossier, tant le sujet est brûlant en matière de cybersécurité. Bien que la connaissance sur le sujet s'avère parfois encore empirique, les experts s'accordent tous à dire que oui, l'IA impactera, et impacte déjà l'univers cyber et est synonyme d'opportunités pour les défenseurs... et pour les attaquants. L'Union européenne elle, tente d'anticiper en adoptant des réglementations plus contraignantes, comme l'AI Act, mais pas que. C'est aussi le très attendu Cyber Resilience Act qui doit globalement venir muscler la protection des produits numériques contre les menaces.

IA et cybersécurité :

une affaire de gentils et de méchants

L'intelligence artificielle (IA) est devenue un outil essentiel dans le domaine de la cybersécurité en raison de sa capacité à analyser de grandes quantités de données en temps réel et à détecter les menaces potentielles de manière proactive. Une nouvelle composante a fait son entrée : l'IA générative, qui trouve déjà des applications du côté des défenseurs... mais aussi des attaquants en transformant le visage de la menace.

L'avènement de l'Intelligence artificielle (IA) ces dernières années soulève de nombreuses problématiques en matière de cybersécurité, d'autant plus depuis l'arrivée des grands modèles de langage (LLM), qui ont rendu l'IA plus accessible que jamais par l'intermédiaire des agents conversationnels. En discutant avec les RSSI (Responsables de la sécurité des systèmes d'information) et les équipes de sécurité des clients de Google, Thiebaut Meyer, directeur stratégies de sécurité chez Google Cloud, a identifié trois grands enjeux en matière de cybersécurité : « *le premier concerne la façon dont l'IA va transformer la menace ; le second, comment elle aidera les équipes de sécurité à être plus performantes et réactives ; le troisième, qui prend de plus en plus d'importance, la façon de déployer une application ou un modèle d'IA de manière sécurisée.* »

La feuille de route est sensiblement identique pour Fred Simon, cofondateur et Chief Data Scientist de JFrog, entreprise spécialisée dans le gestionnaire de dépôts binaires et le DevSecOps. Il identifie lui aussi trois piliers fondamentaux : « *l'IA pour aider les attaquants à former des attaques et à les mettre en place, l'IA pour aider la défense contre les menaces, et l'IA comme nouveau vecteur d'attaques des entreprises.* » Les cybercriminels sont au centre des préoccupations. Et à juste titre, selon Fred Simon : « *nous commençons à observer des cas d'usage de l'IA dont peuvent se saisir les pirates pour écrire du code plus rapidement, par exemple, et essayer de contourner les défenses traditionnelles.* » Bien que la plupart des LLM soient sécurisés pour éviter de produire du code malveillant, nombre de modèles open source ont fait leur entrée sur le marché. « *Une fois qu'il y a possibilité de fournir à une IA du code existant et des modes de défense existants, il est possible d'exploiter ces informations* », ajoute le cofondateur de JFrog. Et ce bien sûr, avec des performances bien supérieures à ce que serait capable de fournir un être humain, en termes de quantité de données traitées et de vitesse d'analyse.

Gen AI : l'ingénierie sociale comme vous ne l'avez jamais vue

Mais là où les observateurs craignent que l'IA, et plus exactement la Gen AI, ait le plus d'impact, du moins dans un premier temps, c'est sur la qualité des attaques d'ingénierie sociale. « *Les attaquants adoptent de plus en plus des stratégies de relations de clients à fournisseurs ; on le voit avec le ransomware en tant que service et la fourniture de kits d'hameçonnage. Ces kits deviennent d'ailleurs bien meilleurs et, dans le fond, ils le sont grâce à l'IA* », souligne Benoît Grunemwald, directeur des relations publiques et porte-parole d'ESET, éditeur de solutions de cybersécurité. Des kits plus efficaces, plus universels, mieux traduits...

La Gen AI permet en fait d'atteindre un haut degré de personnalisation pour rendre des méthodes, tel le spear-phishing (hameçonnage ciblé), toujours plus difficiles à déceler pour une cible. Comment ? En tirant profit des quantités astronomiques de données publiques, afin de dresser un profil toujours plus fin de la cible, en analysant ses comportements, ses préférences, et en produisant ensuite des e-mails de phishing aux petits oignons, et donc en théorie, plus efficaces pour tromper la victime.

Marijus Briedis, CTO de NordVPN, allait même plus loin dans un rapport sur les risques liés à l'association des hackers et de l'IA : « *si un employé d'une entreprise utilise un outil d'IA pour rédiger un rapport à partir d'informations confidentielles, ces mêmes données pourraient être utilisées ultérieurement pour créer des attaques de spear-phishing hautement personnalisées* », avançait dans un rapport sur les risques liés à l'association des hackers et de l'IA. À comprendre,



Eric Antibi, directeur technique chez Palo Alto Networks.



« Sur le long terme, la Gen AI bénéficiera plutôt aux défenseurs qui connaissent l'environnement, les caractéristiques spécifiques de leur organisation et ses données indispensables à l'entraînement des modèles. Cela permettra d'aboutir à des résultats pertinents, au contraire, des attaquants qui disposeront d'outils plus génériques et qui ne seront pas nécessairement adaptés au contexte d'une entreprise ou d'une organisation. »

Thiébaut Meyer,
directeur stratégies de sécurité, Google Cloud.

un employé serait plus crédule face à un e-mail de phishing contenant des informations confidentielles seulement connues de lui et d'une poignée d'autres collaborateurs. Ce dernier exemple soulève d'ailleurs une autre problématique, celle de la mauvaise manipulation et de l'utilisation abusive de données sensibles. Un risque bien réel : dans un rapport de février 2023, l'éditeur de sécurité Cyberhaven avait, en effet, indiqué que 10,8% des employés d'entreprise utilisant sa solution Cyberhaven, avaient utilisé ChatGPT sur leur lieu de travail et 8,6% y avaient renseigné des données d'entreprise, dont certaines sensibles, du code source et des données clients. Les RSSI auront sans doute à cœur de former les collaborateurs aux meilleures pratiques en matière d'utilisation de LLM.

Un risque encore empirique ?

Mais pour l'heure, qu'en est-il vraiment ? Les attaquants se sont-ils saisis de ces outils ? Vincent Nguyen, directeur de la cybersécurité chez Stoïk, agence française de souscription d'assurance spécialisée sur le risque cyber, n'a pas « constaté d'attaques exploitant explicitement de l'intelligence artificielle » sur le portefeuille d'assurés de l'entreprise. « Toutefois, nous utilisons l'IA dans le cadre de certaines de nos activités de cybersécurité et créons à l'occasion des e-mails de phishing (dans le cadre de campagnes de Red Teaming, NDRL). Et

nous savons pertinemment que si nous le faisons, les attaquants le font aussi », prévient Vincent Nguyen. L'expert nuance toutefois son propos, précisant que les attaquants n'ont pas besoin aujourd'hui d'utiliser de fonction avancée d'IA pour mener leurs attaques, dans la mesure où « ils n'ont pas attendu pour automatiser une partie de leur processus de création de Malware ».

Cela dit, si les exemples d'attaques dopées à l'IA ne sont pas légion, les experts s'accordent à dire que l'IA va bel et bien transformer la menace. Au-delà de la crédibilité renforcée des techniques d'ingénierie sociale, « les attaques seront plus véloces », assure Eric Antibi, directeur technique chez Palo Alto Networks. Et déjà, certains signes ne trompent pas. Un rapport de l'équipe de recherche de cybersécurité de Palo Alto Networks, Unit42, s'est penché sur le temps que mettait une attaque perpétrée par les affiliés du fournisseur de ransomware en tant que service Black Basta, pour arriver à ses fins. « Sur des exemples concrets, ce laps de temps était d'environ quatorze heures sans IA », indique Eric Antibi. Une durée qui pourrait être réduite à trois heures en utilisant des outils comme WormGPT. « En temps normal, on estime qu'il s'écoule trente minutes avant qu'un utilisateur ne clique sur un lien de phishing. » Mais une fois corrigées les grossières erreurs de grammaire, de syntaxe et autres tournures de phrase alambiquées, « ce temps est divisé par trois », avance Eric Antibi. Autre exemple édifiant : il s'écoulait en moyenne quarante-quatre jours entre l'intrusion dans un système d'information et le passage à l'action et un vol de données en 2021, un laps de temps descendu à trente jours en 2022... et à cinq en 2023. Eric Antibi craint que ce chiffre ne chute à quelques heures. Et de son avis, « l'automatisation et l'intelligence artificielle font forcément partie de l'équation ». C'est aussi le changement d'échelle qui inquiète le chercheur, les pirates étant désormais en mesure de générer plus rapidement des contenus malveillants de qualité à des échelles industrielles.

Des attaques déjà recensées

Le risque n'est certes pas généralisé, mais les premières attaques dopées à l'IA ont bien été observées. Dans le rapport « Naviguer dans les cybermenaces et renforcer les

Législation de l'IA : les premiers jets

Si les États-Unis sont restés relativement frileux en matière de législation, exception faite d'un décret donnant des orientations, le Parlement européen a adopté l'AI Act, mercredi 13 mars. Il s'agit du premier règlement au monde, véritablement contraignant, encadrant l'intelligence artificielle et ses usages. Le texte doit protéger des risques liés à l'intelligence artificielle (IA) en matière de sécurité et de respect des droits fondamentaux. Le texte contient des dispositions relatives à la protection des données, aux règles sur les données servant à l'entraînement des modèles, entre autres.



Microsoft et OpenAI ont révélé avoir détecté plusieurs groupes de cybercriminels recourant à des LLM pour effectuer des recherches destinées à automatiser ou optimiser leurs techniques de scripting, affiner le ciblage d'organisations, rechercher des vulnérabilités ou encore rédiger des e-mails de spear-phishing.

défenses à l'ère de l'IA », Microsoft et OpenAI ont détecté plusieurs groupes de cybercriminels utilisant des LLM. Forest Blizzard (Strontium), entité associée au renseignement russe, les utilise pour effectuer des recherches afin d'automatiser ou d'optimiser leurs techniques de scripting ; Emerald Street, alias Thallium, pour affiner le ciblage d'organisations, rechercher des vulnérabilités et rédiger des e-mails de spear-phishing ; des hackers nord-coréens pour comprendre les caractéristiques d'une vulnérabilité CVE se rapportant à un outil de diagnostic matériel de Microsoft...

Le risque se cache également dans les modèles open source accessibles aux développeurs. L'équipe de recherche de JFrog a ainsi découvert que des modèles d'IA sur la bibliothèque de modèles open source Hugging Face renfermaient du code malveillant et étaient utilisés comme vecteur d'attaque. Une fois téléchargés sur l'ordinateur du développeur, ils exécutent du code arbitraire, libérant ainsi leur charge utile, accordent un remote shell permettant aux attaques de prendre le contrôle des machines compromises. « Cette infiltration silencieuse pourrait potentiellement donner accès à des systèmes internes critiques et ouvrir la voie à des violations de données à grande échelle, voire à l'espionnage d'entreprise », a prévenu l'équipe de JFrog.

Certains outils sont même ouvertement malveillants. L'exemple le plus connu est le LLM WormGPT. Jumeau maléfique de ChatGPT, il s'appuie sur le modèle open source GPT-J, a été développé pour assister les cybercriminels dans leurs activités. Il passe outre les considérations éthiques et légales imposées à ChatGPT, Bard et consorts. Il peut ainsi répondre à toutes sortes

d'interrogations portant sur des malwares, des arnaques, des codes malveillants, générer des e-mails de phishing.

Les LLM, a priori sécurisés, ne sont pas non plus à l'abri. Sur des forums, des pirates partagent des requêtes et tactiques pour réaliser des attaques de prompt-injection contre les LLM ayant pignon sur rue, et en faire des assistants obéissants. Certaines attaques, dites par empoisonnement, visent à modifier le comportement d'une IA (LLM, ML) en introduisant des données corrompues lors de sa phase d'entraînement et influencer ses prédictions. Ce type d'attaque peut être mené dans le but de neutraliser des systèmes de sécurité, introduire des vulnérabilités, des portes dérobées ou des biais qui pourraient compromettre la sécurité, l'efficacité ou le comportement éthique du modèle. Dans son document « Récit de l'année : impact de l'IA sur la cybersécurité », Kaspersky a d'ailleurs partagé des captures d'écran de cybercriminels sur des forums du Dark Web, partageant du code généré par GPT-4, censé faciliter le traitement des données volées. L'éditeur rappelle toutefois qu'à l'heure actuelle, les performances des LLM sont « limitées », « D'après notre expérience, ils ont tendance à halluciner lorsque les questions et les tâches dépassent un niveau très basique », explique Kaspersky. Toutefois, Thiébaut Meyer nous met en garde : « il est crucial pour les équipes de sécurité de considérer la sécurité de l'IA, de comprendre le fonctionnement des modèles et d'anticiper les risques d'attaques visant à altérer les données d'entraînement. » Les systèmes d'IA, quels qu'ils soient, doivent être testés et mis à jour périodiquement. D'autant plus lorsque l'on est dans un paysage cyber qui évolue rapidement.

Deepfake plus vrai que nature

Contraction de « Deep Learning » et « Fake », les deepfakes opèrent une synthèse de différents médias de manière à générer des trucages possiblement utilisables pour monter arnaques et fraudes. Oracle est revenu dans un article sur la genèse de cette technologie : le GAN, pour Generative Adversarial Networks. Derrière cette définition se cachent deux algorithmes qui s'entraînent et s'améliorent mutuellement. L'un travaille à fabriquer des faux aussi fiables que possible, l'autre tente de les détecter. Ces derniers mois, ces technologies semblent avoir atteint un degré de réalisme jusqu'ici inégalé. « *En ajoutant la voix ou la vidéo, ces deepfakes nouvelle génération permettront de lancer des chaînes d'extorsion plus vraies que nature contre les entreprises privées* », estimait, dans ses prédictions 2024, Vade, société française spécialisée dans la conception et l'édition de solutions logicielles de sécurité des e-mails. Preuve par l'exemple : une récente arnaque en début d'année a coûté 25 millions de dollars à une multinationale hongkongaise. Un employé a été piégé par un e-mail de son directeur lui demandant de rejoindre une visioconférence. Méfiant dans un premier temps, la victime a suspecté une tentative de phishing. Elle a néanmoins été rassurée une fois connectée à la conférence, en voyant plusieurs de ses collègues... qui étaient en fait des deepfakes. Poussé par ses faux collègues, le salarié a effectué une quinzaine de transactions sur cinq comptes bancaires, pour un montant total de 200 millions de dollars de Hong Kong, soit l'équivalent de 25 millions de dollars US.

La défense, une IA bien installée

Les inquiétudes concernant les risques liés à l'IA sont réelles. Cela dit, la technologie est aussi l'affaire de défenseurs et ces derniers n'ont pas attendu l'avènement de la Gen AI pour se l'approprier, notamment à travers le Machine Learning (ML) et être ainsi en mesure de détecter des programmes malveillants. Les applications du ML sont nombreuses en cybersécurité : identification et adaptation des schémas de données pour les rendre plus facilement analysables ; classification des données ; synthèse des possibilités ; identification des profils de sécurité des utilisateurs ; recommandations d'actions de correction...

Dans le cadre d'une remédiation à une cyberattaque, le ML facilite l'extraction des données d'une attaque, leur regroupement, la préparation pour analyse et la fourniture de rapports simplifiés recommandant des actions afin



d'accélérer la prise de décision. Et alors que la plupart des outils dans le monde de la cyberdéfense sont exhaustifs, « *presque 80 % des vulnérabilités qu'ils rapportent ne sont pas applicables pour des raisons techniques* », explique Fred Simon. L'IA fournit plus rapidement, et avec plus d'efficacité, une vision complète de la méthodologie que pourraient employer des cyberattaquants, et certifie de l'applicabilité ou non d'une faille — et donc si celle-ci représente un risque réel et doit faire l'objet d'un patch.

Bien que l'IA soit intégrée depuis des années dans de nombreuses solutions de cybersécurité, « *la maîtrise n'est pas parfaite, beaucoup de systèmes ont des défauts et nous recevons quantité de faux positifs dans nos alertes EDR* », affirme Ayoub M., analyste chez Stoïk. Il s'agit en l'occurrence de bibliothèques suspectes qui, après contrôle, s'avèrent légitimes. Mais si d'aventure ladite bibliothèque embarque des techniques de cryptographie, l'IA peut les confondre avec des modules de chiffrement utilisés par les ransomwares.

Rationaliser l'investigation

Et les LLM dans tout ça ? Ils ajoutent une nouvelle couche d'intelligence dans la défense et renforcent l'efficacité des équipes. Comme expliqué plus haut, « *l'outil ne doit pas se contenter de localiser et lister les CVE (Common Vulnerabilities and Exposures, ndlr), mais doit établir des rapports, lister les actifs exposés en fonction du risque et expliciter les actions correctives à mettre en place* ». À cette fin, Palo Alto Networks a, par exemple, intégré un LLM développé en interne dans sa plateforme de protection des applications cloud-native (CNAPP) Prisma Cloud. À grand renfort de langage naturel, le LLM va lister les actifs exposés les plus à risque, avant de lancer des playbooks automatisés.

Globalement, les grands modèles de langage (LLM) auront, et ont déjà, un impact significatif sur la red team dans l'automatisation des tâches répétitives et fastidieuses, et dans la conduite de campagnes d'ingénierie sociale et de phishing convaincantes et personnalisées à grande échelle, une tâche qui aurait autrefois pris des heures voire des jours. ■

V.M

« *Nous utilisons l'IA dans le cadre de certaines de nos activités de cybersécurité, et nous créons à l'occasion des emails de phishing (dans le cadre de campagne de Red Teaming NDLR). Nous savons pertinemment que si nous le faisons, les attaquants le font aussi.* »

Vincent Nguyen,
directeur de la cybersécurité chez Stoïk.

Checkpoint mise sur l'effet plate-forme

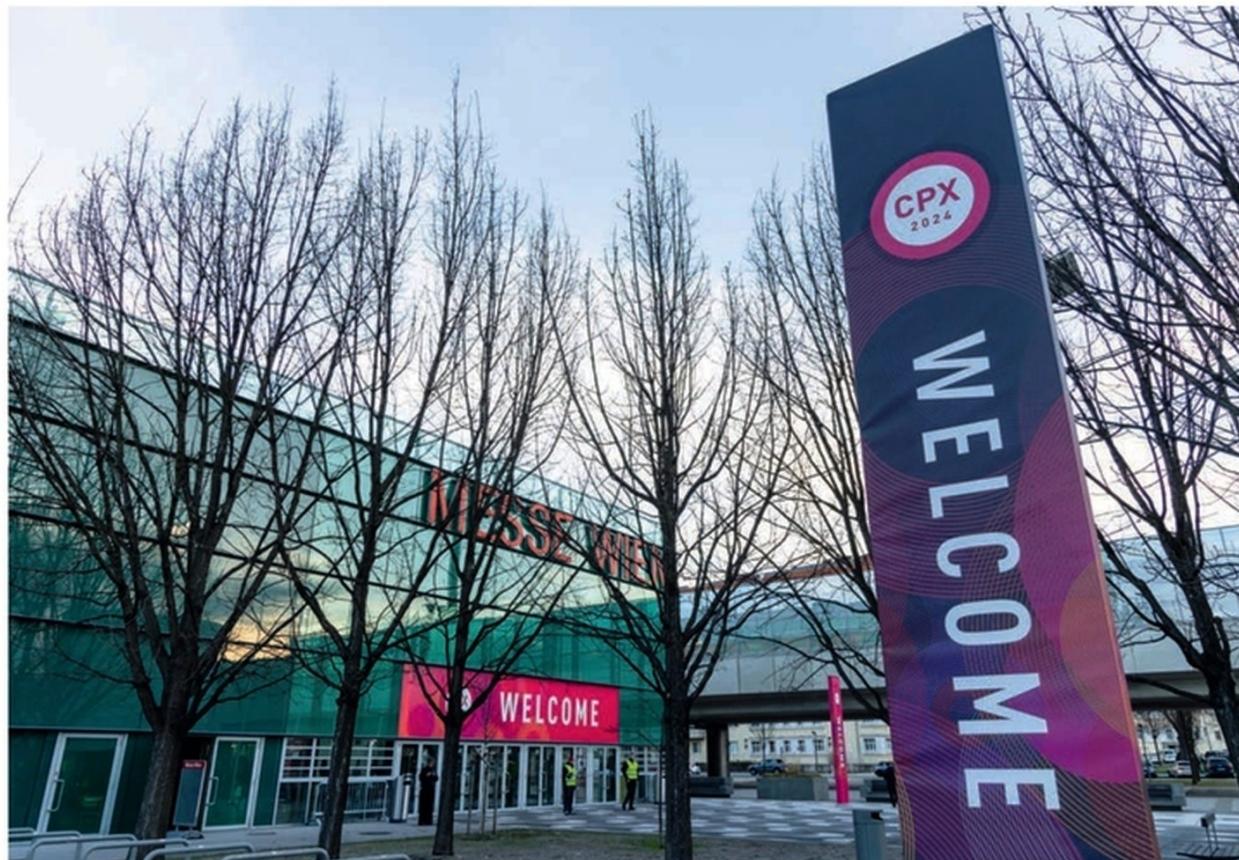
Lors de sa dernière conférence CPX qui s'est tenue à Vienne en Autriche, Gil Schwed, CEO et cofondateur de Checkpoint n'a pas laissé l'ombre d'un doute : le futur de la sécurité informatique se fera par une plate-forme délivrée dans le Cloud utilisant toute les ressources de l'intelligence artificielle.

Comparativement à d'autres acteurs de la sécurité informatique, l'éditeur d'origine israélienne ne multiplie pas les annonces autour de SASE laissant parfois penser qu'il est un peu en retard sur les différentes fonctions nécessaires pour remplir les différentes tâches définies par ce concept créé ex nihilo par le cabinet Gartner en 2018. Mais que ce soit par développement interne ou par croissance externe Checkpoint est désormais à niveau sans avoir à dire que la société est une «SASE Company». Nataly Kremer, Chief Product Officer & Head of R&D, le constate auprès des clients qu'elle

rencontre. «Beaucoup ne discerne pas l'accélération que nous avons effectué et constatent qu'en fait nous avons toutes les fonctionnalités qu'ils recherchent. Notre plate-forme a une très large couverture fonctionnelle et récemment par acquisition nous avons ajouté les briques comme le SD-WAN. Le tout dans une plate-forme unifiée et consolidée sur une seule console qui apporte une expérience utilisateur et des performances qui sont de véritables différenciateurs».

L'IA comme carburant...

Interrogée sur le rôle de l'IA sur la plate-forme, Nataly Kremer indique : «L'IA c'est le carburant de notre plate-forme et sera présente là où la qualité et le coût est le meilleur». D'ailleurs pour elle les cas d'usages se multiplient mais il est nécessaire que le coût de telles solutions baissent encore. Pour y parvenir Checkpoint pré-entraîne les modèles pour assurer performances et qualité des réponses. Lors de l'entretien elle a tenu à rappeler que l'ADN de l'entreprise était la prévention, de résoudre le problème pas de le déplacer vers d'autres. L'aspect collaboratif de la plate-forme entre alors en jeu pour que toutes les parties prenantes dans la sécurité avec l'IA en support pour plus de précision dans la détection et la réponse.





*Gil Schwei,
CEO de Checkpoint
lors de son keynote.*

Et comme assistance intelligente

Infinity Copilot est la traduction concrète du discours de l'entreprise pour l'IA. Cette première génération d'outils mise sur la convergence des technologies de l'IA et du cloud, Infinity AI Copilot répond à la pénurie mondiale croissante de professionnels de la cybersécurité en dynamisant le rendement et l'efficacité des équipes de sécurité. Grâce à la puissance de l'IA générative (GenAI), Infinity AI Copilot assume simultanément les fonctions d'assistant administratif et analytique. La solution automatise les tâches de sécurité complexes et met à disposition des solutions proactives pour lutter contre les problèmes de sécurité. Elle réduit également considérablement le temps consacré aux tâches de routine et permet aux équipes de sécurité de concentrer davantage leur attention sur l'innovation stratégique. Par ailleurs, elle s'intègre parfaitement à la plateforme Check Point Infinity pour offrir une expérience de sécurité unifiée, depuis les terminaux jusqu'au réseau, en passant par le cloud.

La solution prend en charge la gestion, les modifications et les déploiements automatique des règles d'accès et les contrôles de sécurité spécifiques à la politique de chaque client. Infinity AI Copilot permet de superviser tous les produits de l'ensemble de la plateforme Check Point Infinity, du réseau au cloud en passant par l'espace de travail, pour en faire un véritable assistant polyvalent. les échanges avec Infinity AI Copilot GenAI sont aussi simples qu'avec un être humain. Le système comprend et répond via un chat accessible dans toutes les langues pour simplifier la communication et l'exécution des tâches pour les utilisateurs. Le langage naturel permet

une interaction transparente et une exécution efficace des tâches. Selon l'éditeur l'assistant permet d'économiser jusqu'à 90% du temps passé à la gestion des tâches de sécurité, comme l'analyse des événements, les mises en œuvre et le dépannage. Infinity AI Copilot est actuellement disponible dans sa version préliminaire. Le lancement de la version intégrale quant à elle est prévue au deuxième trimestre. Les futures innovations comprennent des fonctions d'assistance proactive et de gestion autonome des politiques. Ces dernières fonctions devraient accompagnés des annonces majeures au mois de juin prochain.

Les PME ne sont pas oubliées

Lors de l'événement Checkpoint a aussi annoncé deux nouvelles familles de pares-feux de nouvelle génération, les Quantum Spark 1900 et 2000 à destination des PME, selon la définition de Checkpoint. Les appareils proposent un système de sécurité réseau performant à l'échelle, grâce à un dimensionnement sans intervention, une gestion avancée du cloud et une gestion automatisée des menaces. Ces passerelles sont donc parfaitement adaptées aux PME. La gamme complète de passerelles permet de répondre

aux besoins des entreprises de 1 à 1000 employés. Elles offrent aux PME un outil qui combine la prévention des menaces par l'IA, une gestion facilitée, et une capacité d'évolution. Un atout d'autant plus important que de nombreuses PME sont confrontées à des ressources et à une expertise limitées en matière de cybersécurité. Le lancement de ces nouveaux produits permettra aux PME de mieux protéger leurs opérations en dépit de la complexité de l'environnement actuel des cybermenaces. Selon Checkpoint les passerelles connaissent un taux de blocage impressionnant de 99,7% contre les attaques de phishing, de malware, de DNS et d'IoT. Elle permet une meilleure résilience, plus d'efficacité et une flexibilité du réseau grâce à des commutateurs intégrés et une double alimentation qui éliminent le besoin de commutateurs externes et maximisent les opérations en minimisant les temps d'arrêt. Les Quantum Spark embarquent des services intégrés de sécurité dans le Cloud et de connectivité réseau. Des fonctions de gestion du Cloud sont aussi présentes pour simplifier le travail des fournisseurs de services de sécurité (MSP) avec un centre centralisé pour les services clients, les licences et la surveillance proactive de la sécurité, accessible via un tableau de bord unifié. Ces fournisseurs de services devraient représenter 40 % des dépenses en sécurité en 2025 selon les prévisions.

Les séries Quantum Spark 1900 et 2000 sont actuellement disponibles. Pour plus d'informations et en savoir plus sur les caractéristiques des six modèles de la gamme Quantum Spark Gateway, rendez-vous sur le site : <https://www.checkpoint.com/quantum/next-generation-firewall/small-business-firewall/> ■

B.G

Chez Gatewatcher, nous vous donnons *le pouvoir de prévenir les menaces* afin que vous puissiez protéger vos données et travailler en toute sécurité.

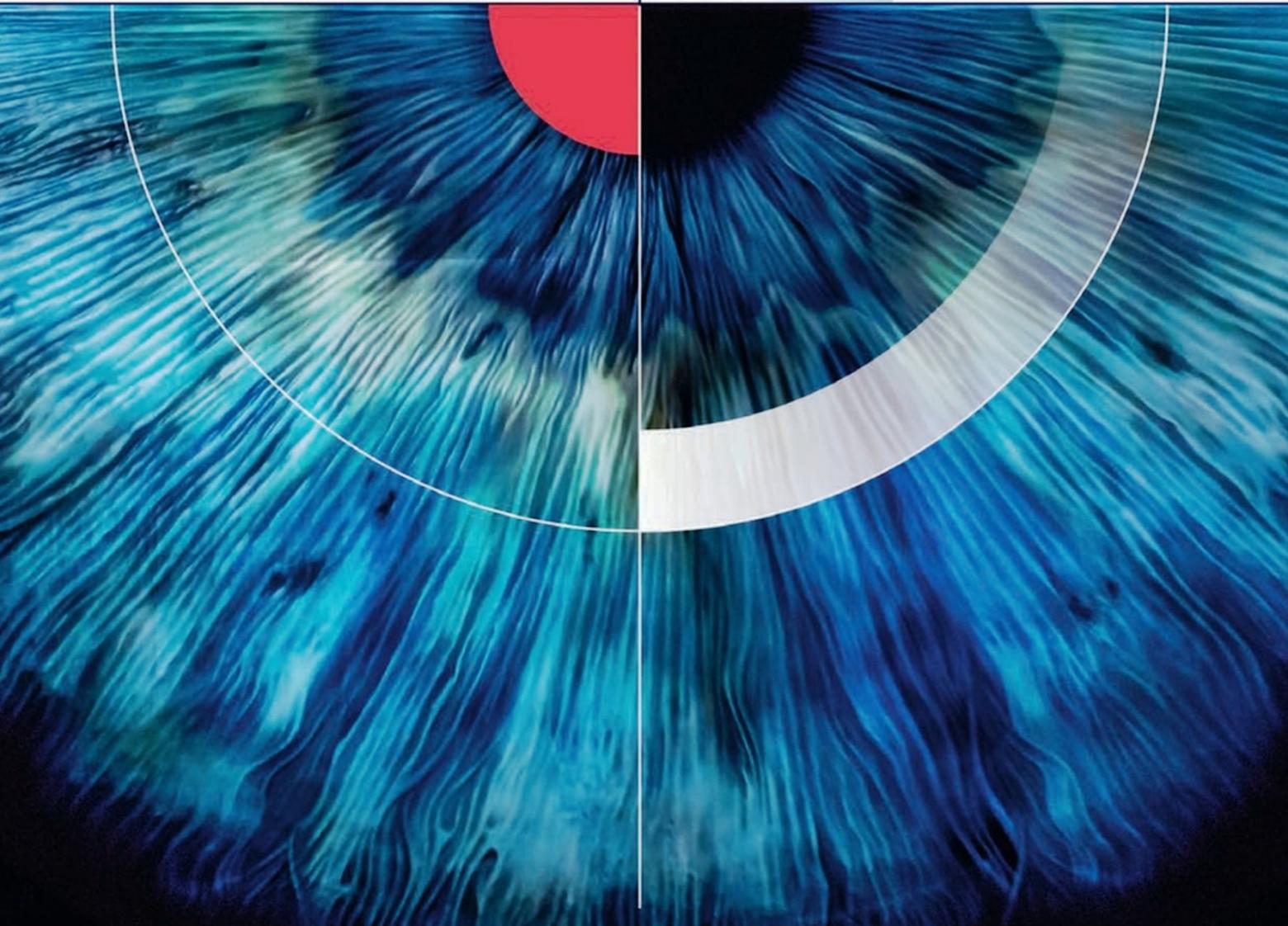
Notre *objectif* : vous offrir jour après jour la tranquillité d'esprit dont vous avez besoin pour vous concentrer sur votre activité.

+

gatewatcher.com



-  **NDR**
-  **CTI**
-  **SENSORS**
-  **TAP**



AFCDP revient sur **20 ans** de Protection des Données Personnelles

Paul-Olivier Gibert, Président de l'AFCDP.

Créée en 2004, l'association française AFCDP (Association Française des Correspondants à la Protection des Données à caractère Personnel) a pour rôle principal de promouvoir et de défendre les droits liés à la protection des données à caractère personnel.

L'aventure commence avec la directive de 1995, suite au texte de la loi Informatique et Libertés de 1978 en France, qui constituait l'un des premiers textes législatifs européens définissant les droits des individus en matière de traitement des données personnelles. Cette directive, qui avait prévu la possibilité de désigner des correspondants à la protection des données, a été transcrit dans une évolution de la Loi Informatique et Libertés en 2004, avec l'instauration du correspondant à la protection des données, ou CIL (correspondant informatique et libertés). La création de l'AFCDP en 2004 a répondu au besoin de mettre en place une structure regroupant les CIL et les professionnels spécialisés dans la protection des données. Son but était d'accompagner les organisations dans la mise en œuvre de bonnes pratiques en la matière, conformément aux évolutions législatives et réglementaires, nationales et européennes. Puis, le métier de DPO (Data Protection Officer), ou Délégué à la Protection des Données, a été créé par le Règlement Général sur la Protection des Données (RGPD), réglementation de l'Union européenne entrée en vigueur en mai 2018.

À travers plusieurs outils, l'AFCDP offre à tous les professionnels soucieux de la protection des données personnelles un espace de partage d'expériences, de connaissances et de bonnes pratiques entre pairs, chargés de veiller à la conformité et au respect de la législation sur la protection des données au sein des organisations. Le plus emblématique de ces outils étant sa plateforme d'échanges qui permet à tout et à chacun d'exposer une problématique et d'y voir répondre des experts du domaine (DPO, avocats, informaticiens, ...).

Ainsi, depuis 18 ans, l'association réunit ses membres annuellement en début d'année pour l'Université des DPO, qui se décline tout au long de l'année avec des rendez-vous mensuels sur les enjeux techniques, métiers et marchés que rencontrent les professionnels au quotidien.

... au-delà des frontières de l'écosystème

Ces enjeux étant transverses à tous les secteurs, l'association a souhaité créer des groupes de travail thématiques (Marketing, Santé, Droit Européen, etc.), en plus des groupes régionaux existants partout sur le territoire, y compris en outre-mer.

Au fil des années, l'AFCDP, partenaire de la CNIL et de Cybermalveillance.gouv.fr notamment, a été un acteur actif dans les débats et les évolutions législatives en matière de protection des données en France et en Europe. Elle a contribué aux discussions sur le règlement général sur la protection des données (RGPD), qui est entré en vigueur en 2018 et qui a renforcé les droits des individus et les obligations des organisations en matière de traitement des données personnelles.

L'évolution rapide des technologies de l'information et la globalisation des échanges ont conduit à une prise de conscience croissante de l'importance de protéger la vie privée des individus, et l'AFCDP continue d'œuvrer dans ce sens dans les événements professionnels auxquels elle participe, mais aussi par voie de presse notamment.

Aujourd'hui, l'association va encore plus loin en représentant les intérêts et les préoccupations de ses 6000 membres auprès des législateurs. L'objectif étant qu'ils soient pris en compte de manière adéquate dans l'élaboration de toutes les politiques numériques, assurant ainsi une meilleure harmonisation entre les exigences réglementaires et les réalités pratiques du terrain.

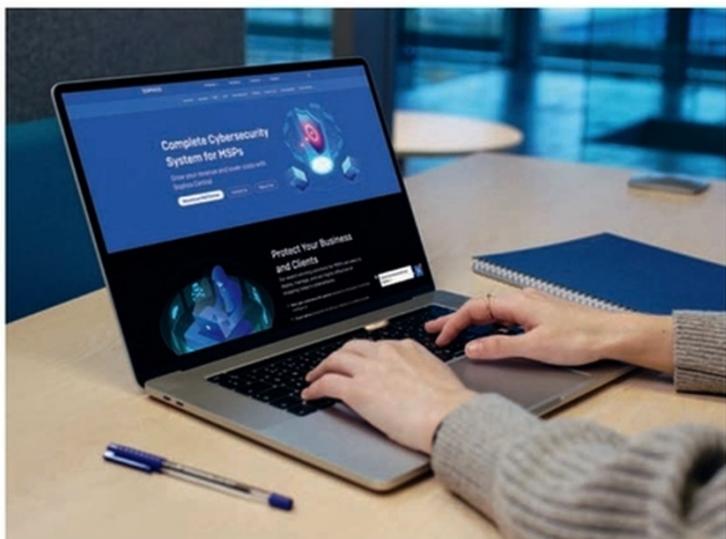
En résumé, l'AFCDP joue un rôle crucial dans le paysage de la protection des données en France, et en Europe au travers de la CEPDO (dont l'association est membre fondateur), en rassemblant des professionnels engagés dans la promotion et la défense de la vie privée et des droits individuels, dans un contexte numérique et législatif qui ne cesse d'évoluer, et qui promet une accélération des challenges à relever dans les 20 prochaines années. ■



Étude de Sophos : décryptage de l'utilisation réelle ou supposée de l'IA par les cybercriminels

Le spécialiste de la sécurité informatique « as a Service », Sophos, a récemment publié deux études sur l'utilisation potentielle de l'IA par les cybercriminels. Celles-ci révèlent une situation plus complexe qu'il n'y paraît. Si certains utilisent déjà l'IA pour parfaire leurs escroqueries, d'autres se montrent beaucoup plus sceptiques vis-à-vis de ces nouvelles technologies.

L'étude « *The Dark Side of AI : Large-Scale Scam Campaigns Made Possible by Generative AI* » a pour but d'explorer comment les technologies génératives comme ChatGPT d'OpenAI pourraient permettre aux cybercriminels de fomenter des fraudes massives. En utilisant des templates simples d'e-commerce et des outils LLM (Large Language Model) comme GPT-4, les experts de Sophos X-Ops (une équipe réunissant différentes divisions de la société) a réussi à créer un site web de commerce entièrement fonctionnel, avec des images, des messages audio, et des descriptions de produits générés par IA. Ils ont complété le tout par de fausses pages de connexion via Facebook et de paiement destinées à dérober les identifiants et les coordonnées bancaires des visiteurs.



De l'automatisation aux LLM

Cette démonstration prouve, s'il en est besoin, à quel point il est aisément pour les cyberescrocs de mettre sur pied un site web frauduleux avec un minimum de connaissances techniques. Sophos X-Ops a ensuite créé des centaines de sites web similaires en seulement... quelques minutes ! « *Il est naturel — et attendu — que les cybercriminels se tournent vers de nouvelles technologies pour automatiser leurs attaques. La création des premiers e-mails de spam avait constitué une étape cruciale en faisant franchir aux techniques d'escroquerie un nouveau palier en termes d'échelle. De nouvelles IA s'apparentent à faire de même : s'il existe une technologie d'IA capable de créer des menaces entièrement automatisées, elle finira par être exploitée à cette fin. Nous avons d'ailleurs déjà observé l'intégration de textes ou de photos générés par IA pour servir d'appâts dans des escroqueries classiques* », commente Ben Gelman, Senior Data Scientist chez Sophos.

Anticiper les cyberattaques basées sur l'IA

La seconde étude « *Cybercriminals Can't Agree on GPTs* » menée par les experts met à mal certaines idées reçues selon lesquelles les cybercriminels utiliseraient massivement les IA génératives pour développer des charges malveillantes. En infiltrant les principaux forums du dark web, ils se sont rendus compte que l'IA dans la cybercriminalité n'en était qu'à ses balbutiements. Si les cybercriminels sont bien conscients du potentiel des LLM pour l'ingénierie sociale, ils se montrent beaucoup plus sceptiques quant à leur utilisation pour la création de malwares. « *Tandis que d'importantes préoccupations se sont fait jour concernant l'utilisation abusive de l'IA et des LLM par des cybercriminels depuis le lancement de ChatGPT, notre étude révèle que, jusqu'ici, les acteurs malveillants font preuve de plus de scepticisme que d'enthousiasme. Sur deux des quatre forums que nous avons parcourus sur le Dark Web, nous n'avons qu'une centaine de messages parlant de l'IA, contre un millier à propos des cryptomonnaies au cours de la même période. Nous avons bien vu des cybercriminels tenter de créer des malwares ou des outils d'attaque au moyen de LLM mais les résultats sont rudimentaires et souvent accueillis avec scepticisme par les autres utilisateurs. Pour l'instant, il semble que les cybercriminels nourrissent les mêmes débats sur les LLM que le reste de la population* », ajoute Christopher Budd, directeur de recherche pour Sophos X-Ops. ■

« Faire face aux menaces émergeantes dans le domaine de la sécurité informatique est plus crucial que jamais »

John Shier-Field CTO Commercial chez Sophos.

John Shier, Field CTO Commercial chez Sophos revient sur les points importants que révèlent ces deux études.

Pouvez-vous nous expliquer ce qui a motivé Sophos à mener cette étude sur l'utilisation de l'IA par les cybercriminels ?

Il y a deux raisons. La première, c'est qu'en tant que spécialiste de la cybersécurité, on doit pouvoir prévoir comment les cybercriminels vont potentiellement utiliser l'IA et où ils en sont actuellement dans son développement. La seconde, c'est comment nous pouvons utiliser l'IA nous-même pour que nos produits et nos services soient plus performants pour nos clients. Pour réaliser ces études, nous avons infiltré les principaux forums du dark web et les cercles privés des cybercriminels pour essayer de savoir ce qu'ils disent sur l'IA et comment ils la perçoivent.

D'après les informations que vous avez obtenues, les cyberattaques basées sur l'IA deviennent-elles une réalité ?

Des experts de Sophos avaient déjà mené une autre étude, il y a quelques mois, pour savoir s'il serait possible d'utiliser les outils basés sur l'IA accessibles au public pour effectuer des cyberattaques à grande échelle. Ils ont découvert que c'est effectivement possible, mais à condition de posséder certaines compétences techniques complémentaires. Indépendamment de notre étude, on a pu constater que les cybercriminels ont certainement utilisé l'IA pour des arnaques financières, mais pas pour créer des logiciels malveillants comme des malwares, des ransomwares, etc. Ils exploitent le potentiel de l'IA pour faire de l'ingénierie sociale, et du phishing. Cela leur permet notamment de créer des campagnes de phishing réalistes de plus en plus parfaites et dans différentes langues étrangères. Ils ne sont pas capables pour le moment de créer des malwares uniquement avec une IA générative. Pour accéder aux fonctions avancées et contourner les protections d'un service comme ChatGPT, il faut un certain nombre de « tokens » (jetons). Le développement d'un malware fiable avec une IA générative nécessiterait tellement de tokens qu'ils les épuiseraient avant de pouvoir parvenir à un résultat satisfaisant. Les alternatives open source de ChatGPT ne sont, quant à elles, pas assez puissantes pour créer des logiciels malveillants. Dans une certaine mesure, les IA génératives peuvent permettre de créer des sites web frauduleux. Mais là encore, il y a des limitations et des barrières techniques qui nécessitent de maîtriser d'autres compétences.



MENACES

83

L'étude met en lumière le scepticisme des cybercriminels à l'égard des IA génératives. Avez-vous identifié des avantages ou des inconvénients spécifiques ?

Le plus grand inconvénient pour eux, c'est la difficulté de contourner les barrières techniques des IA génératives, et obtenir un résultat plus fiable ou même égal à ce qu'ils ont aujourd'hui. C'est la raison pour laquelle, selon moi, on ne voit pas beaucoup de logiciels malveillants créés avec l'IA. Aujourd'hui, les techniques et les stratégies qu'ils emploient pour contourner nos systèmes, nos réseaux et pour créer des charges malveillantes fonctionnent très bien. Ils ont déjà beaucoup de succès (on a vu que les ransomwares ont rapporté plus de 1 milliard de dollars en 2023), donc ils n'ont aucune raison de faire plus d'efforts pour intégrer des IA.

PROPOS RECUEILLIS
PAR JÉRÔME CARTEGINI



Converged SIEM

Accélérer la détection, l'investigation et la réponse aux menaces sur une plateforme unique.



Logpoint.com

Le Cyber Resilience Act, la loi de l'UE sur la cybersécurité

De nouvelles règles en matière de cybersécurité ont été édictées par l'Union européenne afin de garantir la sécurité du matériel et des logiciels et, de fait, des utilisateurs, particuliers ou organismes. Cet article va explorer les aspects couverts par ces réglementations.

Les composants numériques sont omniprésents dans notre vie quotidienne. Nombre de produits en contiennent, des jouets pour enfants jusqu'aux voitures en passant par la domotique et autres produits du quotidien. Dès qu'il y a un logiciel et une connexion Internet, même non permanente, il y a un risque en matière de sécurité. La CRA (loi sur la cyber-résilience) vise justement à protéger les consommateurs et les entreprises qui possèdent et utilisent des produits et/ou des logiciels avec un composant numérique, quel qu'il soit. Cette loi introduit des exigences obligatoires en matière de cybersécurité pour les fabricants et détaillants d'objets connectés, avec une protection qui s'étend tout au long du cycle de vie. Elle couvre un champ très large allant des produits grand public aux objets de la « smart city » tels que les compteurs intelligents de type Linky, les hyperviseurs et autres caméras connectées. Il concerne aussi les éléments clefs des systèmes d'information que sont les antivirus, les pare-feux, les routeurs et autres composants hardware et software réseaux. Ces produits seront cependant associés à différents niveaux de risque d'exigence. La problématique abordée par ce règlement est double. En premier lieu, le niveau insuffisant en termes de cybersécurité inhérent à la plupart de ces produits et/ou de leurs mises à jour logicielles. Deuxièmement, à l'heure actuelle, les consommateurs et les entreprises sont le plus souvent incapables de déterminer quels produits sont, ou non, sécurisés, et de mettre en place des protocoles destinés à garantir cette cybersécurité. La loi sur la cyber-résilience est censée garantir plusieurs choses, dont notamment :

- des règles harmonisées pour la mise sur le marché des produits ou logiciels comportant un composant numérique ;

- un cadre d'exigences en matière de cybersécurité régissant à la fois la planification, la conception, le développement et la maintenance de ces produits, avec des obligations à respecter à chaque étape de la chaîne de valeur ;

- l'obligation d'assurer le devoir de sollicitude pour l'ensemble du cycle de vie de ces produits.

Dès que ce règlement entrera en vigueur, les logiciels et tous les produits connectés à Internet porteront un marquage CE afin d'indiquer s'ils sont, ou non, conformes à ces nouvelles normes. Le fait d'exiger des fabricants et des détaillants, qu'ils donnent la priorité à la cybersécurité, permet aux clients et aux entreprises de faire des choix plus éclairés et, sans avoir à vérifier systématiquement ce qu'il en est, d'être confiants dans les qualifications des produits marqués CE. Le CRA ne date pas d'hier, mais plutôt d'avant-hier. Il avait déjà été annoncé dans la stratégie 2020 de l'UE en matière de cybersécurité. Il complète d'autres législations dans le même domaine comme particulièrement le cadre SRI2. Il s'appliquera donc à tous les produits connectés directement ou indirectement à un autre dispositif ou réseau, à l'exception de certaines exclusions spécifiées. Celles-ci concernent notamment les logiciels ou services open source déjà couverts par des règles existantes jugées satisfaisantes comme les dispositifs médicaux, l'aviation et les véhicules de transport. Le règlement devrait entrer en vigueur très prochainement. Les fabricants seront obligés d'appliquer les règles édictées au plus tard 36 mois après leur entrée en vigueur sous peine de sanctions. La Commission européenne examinera ensuite périodiquement le bon fonctionnement de la Loi et fera des rapports en vue de l'améliorer progressivement.



Le CRA, Cyber Resilience Act, vient enfin poser les fondations au niveau de l'Union européenne de la sécurité des objets connectés et de ceux qui les utilisent.

Publications Presse Contact Déclaration vulnérabilité Rejoignez-nous Incident EN

Republique Française
Liberé Egalité Fraternité

ANSSI

Rechercher...

Découvrir l'ANSSI Découvrir la cybersécurité Développer des solutions de confiance Sécuriser son organisation Se former à la cybersécurité Connaitre et explorer S'informer sur la réglementation

Accueil > S'informer sur la réglementation > La confiance numérique > Le cadre de certification de cybersécurité européen > Cybersecurity Act

Cybersecurity Act
Règlement européen « Cybersecurity Act » relatif à l'ENISA et à la certification de cybersécurité des technologies de l'information et des communications.

86

Publié le 18 Août 2022 • Mis à jour le 13 Novembre 2023
Adopté par le Parlement européen le 12 mars puis par le Conseil de l'Union européenne le 17 avril 2019, le

Si vous souhaitez lire dans son intégralité le Cybersecurity Act en français, rendez-vous à l'adresse <https://eur-lex.europa.eu/legal-content/fr/TXT/PDF/?uri=CELEX:32019R0881&from=EN>.

Sinon, vous en trouverez un bon résumé et différents liens utiles sur le site du gouvernement dédié à la sécurité, à l'adresse <https://cyber.gouv.fr/cybersecurity-act-a-description>.

Mise à jour de sécurité

Ce règlement exige clairement des fabricants qu'ils traitent les failles de sécurité importantes, et ce, tout au long du cycle de vie de leurs produits, et qu'ils signalent dans les plus brefs délais toute faille activement exploitée ou pouvant l'être aux autorités jugées compétentes. Ils devront notamment fournir des mises à jour de sécurité pendant toute la durée de vie des produits. La gestion de ces informations de sécurité reviendra aux Centres de réponses aux incidents de sécurité informatique (CSIRT) nationaux en lien avec l'Agence européenne pour la cybersécurité (ENISA) pour les risques systémiques (l'ANSSI en ce qui concerne la France). Les produits développés exclusivement pour la

sécurité nationale ou la défense seront exclus de facto du champ d'application de ce règlement. Concernant les logiciels libres, l'accord stipule que seuls les logiciels développés dans le cadre d'activités commerciales seront concernés par la réglementation. Les responsabilités en matière de documentation et de gestion des failles de sécurité seront donc amoindries pour les développeurs de logiciels libres. Les organisations à but non-lucratif qui distribuent des logiciels libres, mais réinvestissent tous les revenus dans des activités non-lucratives seront exclues du champ d'application direct de la législation. Cette distinction est importante, car elle vise à protéger les communautés de logiciels libres tout en assurant la sécurité des produits commerciaux susceptibles d'intégrer ces logiciels.

Qui est concerné par le CRA ?

La loi concerne en premier lieu les fabricants et fournisseurs de produits, services et processus des technologies de l'information et de la communication (TIC) afin de leur donner un cadre en matière d'exigences de sécurité. Elle concerne, bien entendu, les organismes d'évaluation de la conformité en charge de délivrer des certificats de cybersécurité européens qui devront avoir recours à des méthodes d'évaluation robustes et harmonisées. Enfin, elle concerne évidemment les utilisateurs finaux et les donneurs d'ordre, leur permettant d'utiliser et de choisir des produits, services et processus TIC correspondant à leurs besoins en termes de sécurité.

Entrée en vigueur et champ d'application

L'accord conclu est désormais soumis à l'approbation formelle du Parlement et du Conseil européen. Une fois adopté, le Cyber Résilience Act entrera en vigueur précisément le 20e jour suivant sa publication au Journal officiel. Le CRA étant un acte juridique européen de portée générale, il est obligatoire dans toutes ses dispositions. Les États membres seront donc tenus de les appliquer telles qu'elles sont définies par le règlement. Comme il s'agit d'un règlement d'application directe, les États membres devront mettre leur organisation nationale en conformité avec les dispositions du règlement au maximum dans les deux années suivant sa publication. ■

Chiffrement homomorphe

Une méthode de cryptographie de plus en plus crédible

Le chiffrement entièrement homomorphe (FHE), mode de cryptographie permettant des calculs sur des données chiffrées sans qu'il ne soit nécessaire d'accéder à leur contenu en clair, suscite de plus en plus d'intérêt. Longtemps jugée comme un fantasme dans le milieu des cryptographes, cette technologie est désormais sérieusement envisagée pour diverses applications dans des domaines traitant des informations sensibles.

Fer de lance de la protection des données personnelles et de la vie privée, la cryptographie semble voir émerger le « boss final » dans le domaine avec le chiffrement complètement homomorphe, ou FHE (Fully Homomorphic Encryption). Cette méthode de chiffrement autorise des calculs complexes sur des données sans les déchiffrer, offrant ainsi une sécurité accrue tout au long du processus de traitement. Seul le résultat peut être décrypté, préservant ainsi la confidentialité de la data à toutes les étapes.

Un démarrage laborieux

Le concept du FHE a été théorisé pour la première fois dans les années 1970, mais il n'était pas considéré comme

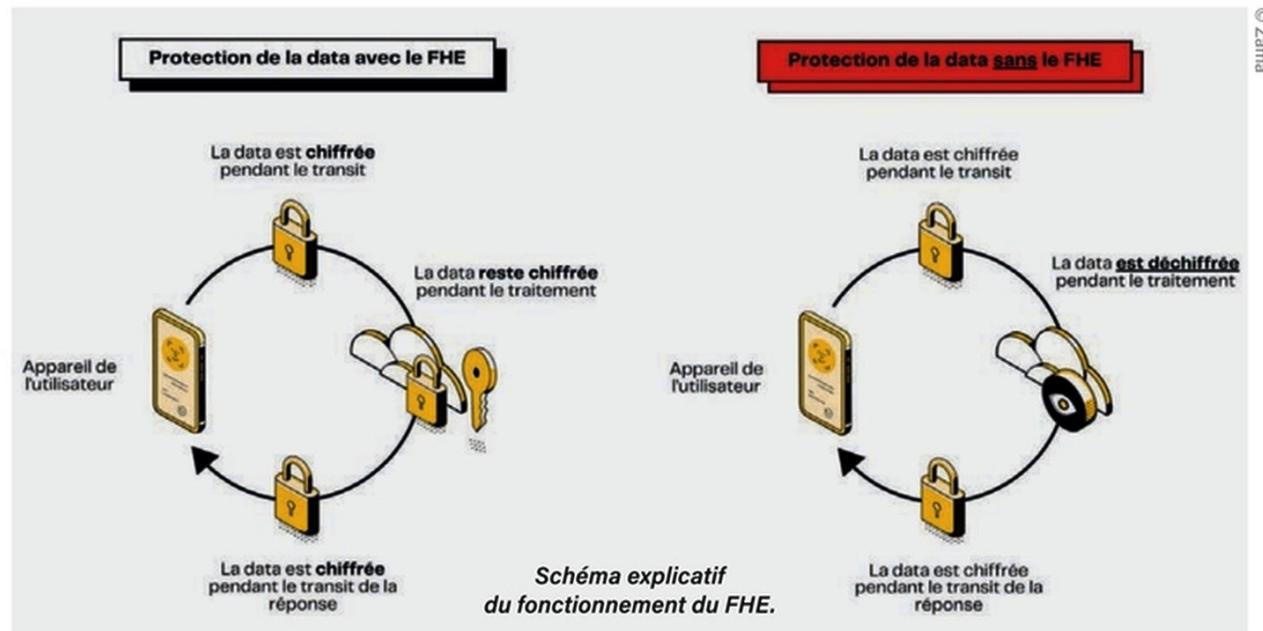


Craig Gentry est le premier à avoir développé un protocole cryptographique homomorphe fonctionnel.

une solution technologique viable, malgré ses promesses. « *Ce fut un rêve pendant des années, car nous ne savions pas comment le réaliser techniquement* », explique Pascal Paillier, cryptographe de formation, cofondateur et CTO de Zama, société de cryptographie open source qui développe des solutions FHE pour la blockchain et l'IA. Ce n'est qu'en 2009 que les choses ont commencé à évoluer grâce à un informaticien américain, Craig Gentry, qui développa le premier protocole cryptographique homomorphe fonctionnel. « *C'était la première fois qu'un mécanisme confirmait l'existence d'un schéma supportant toutes les opérations du FHE. Cependant, il était extrêmement inefficace en raison des temps de latence rédhibitoires* », précise Pascal Paillier.

Pascal Paillier est le CTO et cofondateur avec Rand Hindi (PDG) de Zama, startup spécialisée dans la cryptographie open source et homomorphe.





Toutefois, avec l'amélioration continue des algorithmes et de la puissance de calcul, l'applicabilité du FHE s'est améliorée, perfectionnant ainsi le traitement de grands volumes de données avec des latences de plus en plus faibles. D'où une crédibilité accrue de cette technologie.

En attestent les levées de fonds réalisées par des startups spécialisées dans ce domaine. Citons les 30 millions de dollars levés par Duality Technologies en 2021 et les 4,2 millions d'euros par Cosmian en 2022. Quant à Ravel Technologies, l'un des leaders du secteur, lui aussi a réalisé une levée de fonds cette même année, mais son montant n'a pas été révélé. Plus récemment, début mars 2024, Zama, la société co-fondée par Pascal Paillier et Rand Hindi, PDG, a annoncé avoir levé 73 millions de dollars en Série A, dans une opération dirigée par Multicoin Capital et Protocol Labs, avec la participation de Metaplanet, Blockchange Ventures, VSquared, Stake Capital, Juan Benet (fondateur de Filecoin), Anatoly Yakovenko (co-fondateur de Solana) et Gavin Wood (cofondateur d' Ethereum et cocréateur de Polkadot). Avec ces fonds, la jeune pousse souhaite renforcer son équipe de 75 personnes en recrutant de nouveaux ingénieurs, développeurs logiciels et chercheurs en cryptographie. L'ambition finale étant de créer une nouvelle classe d'applications de chiffrement entièrement homomorphes (FHE) pour les développeurs.

Des opportunités multiples

Les investisseurs, en grande partie issus de l'univers de la blockchain et du Web3, voient, semble-t-il, un fort potentiel dans cette technologie. Dans un écosystème où la vie privée est essentielle, mais où les transactions financières sont visibles, le FHE peut offrir des perspectives intéressantes. Avec la multiplication des applications décentralisées, le besoin de technologies de chiffrement garantissant la confidentialité des données et des transactions devient crucial. Les cas d'utilisation du FHE s'appliquent également dans d'autres secteurs traditionnels. Par exemple, dans le domaine de la publicité ciblée, la santé ou la finance, le FHE peut garantir la confidentialité des données et assurer la conformité réglementaire.

Zama s'est d'abord concentré sur le développement de smart contracts (contrats intelligents, ndlr) confidentiels sur la blockchain et dans le monde du Web3, mais les cas d'utilisation du FHE peuvent s'appliquer à d'autres écosystèmes. Précision d'importance : pour que le FHE soit largement adopté, il est essentiel de fournir des outils clés en main aux développeurs. De plus, le FHE reste gourmand en calcul, ce qui constitue un défi matériel pour son déploiement à grande échelle. À cet égard, le développement d'accélérateurs matériels spécialisés est envisagé pour en améliorer les performances. ■

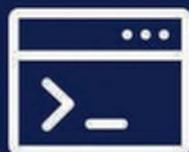
V.M

Une histoire de hardware ?

Pour les petites opérations de transaction en FHE sur la blockchain, l'utilisation d'accélérateurs matériels n'est pas indispensable. « *Le temps de calcul avec le CPU sera de quelques secondes, ce qui suffira pour la blockchain. En revanche, pour une expérience en temps réel sur un grand modèle de langage (LLM), ces accélérateurs seront nécessaires* », indique Pascal Paillier. Les cartes matérielles spécialisées peuvent être intégrées aux serveurs d'un centre de données pour augmenter la puissance de calcul. « *Nous travaillons en étroite collaboration avec les sociétés qui produisent ces technologies pour adapter nos solutions à leur architecture matérielle.* » Pascal Paillier estime que cela permettrait de multiplier par dix mille la vitesse d'exécution. C'est impératif pour atteindre ce que le CTO de Zama appelle « *la suprématie homomorphe* », c'est-à-dire le moment où, grâce à l'accélération matérielle, il deviendra plus avantageux de chiffrer par défaut l'ensemble des données, car il n'y aura plus de différence de temps d'exécution entre ce qui est chiffré et ce qui ne l'est pas. Un objectif (ou graal) qui soulève de nouveaux défis, notamment en matière de consommation d'énergie supplémentaire et d'optimisation.



BACK UP AND KEEP CALM



Operate



Secure



Protect

Leader français de la protection des données



Contact
www.antemeta.fr
+33 1 85 40 03 36

AntemetA accompagne les directions dans la sanctuarisation et l'évolution de leur Système d'Information.

AntemetA, tiers de confiance, assure le plan de reprise d'activité en cas de cyberattaque par la mise en œuvre en amont de solutions d'infrastructure, la fourniture de services Cloud et une expertise des services managés.



Gartner

HEXATRUST
CLOUD CONFIDENCE & CYBERSECURITY



« Un bon pentester doit savoir vulgariser »

Guillaume Lopes, Senior Penetration Tester et CTO chez RandoriSec.

Le cabinet de conseil français Wavestone a présenté, en avril 2023, son baromètre annuel sur la maturité cyber en France. Il en ressortait que le niveau de maturité des grandes organisations était encore faible (49 %), mais en amélioration de trois points par rapport à l'année précédente. Comment tirer ces statistiques vers le haut ?

De l'avis de Guillaume Lopes, Senior Penetration Tester et CTO chez RandoriSec, le pentest, bien qu'il ne soit pas exhaustif, est une bonne entrée en matière pour éléver son niveau de cybersécurité.

L'InfoCR : Quelles sont les qualités qu'un pentester doit réunir pour répondre aux attentes d'un client ?

Le client attend du pentester qu'il soit démonstratif, car l'objectif est de savoir comment un acteur malveillant pourrait l'attaquer, que ce soit sur une application web, un réseau interne ou Wifi. Une des qualités principales d'un pentester est sa capacité à démontrer concrètement comment un attaquant est en mesure d'exploiter une vulnérabilité, comment il est en mesure d'accéder à des données sensibles et de rebondir sur d'autres actifs du système d'information. C'est beaucoup plus probant pour le client, d'autant plus pour ceux qui manquent de maturité cyber. Pour toutes ces raisons, le pentest est une bonne entrée en matière pour un client qui souhaite éléver rapidement son niveau de cybersécurité.

Néanmoins, cette méthode n'est pas exhaustive. Il faut bien faire le distinguo entre le test d'intrusion et l'audit de sécurité, car la démarche est complètement différente. Le pentest constitue une image à un instant donné des vulnérabilités présentes sur une application ou un système d'information. Or, le SI est un écosystème vivant qui évolue très vite. C'est pourquoi, il faudra approfondir avec un audit complémentaire. Cependant, nous encouragerons toujours un client n'ayant jamais réalisé d'audit de sécurité à débuter par un pentest en raison de son aspect démonstratif. Il est vrai que souvent la liste des actions correctrices d'un audit de sécurité peut être déroutante pour un client.

Les entreprises ont-elles conscience de l'importance du pentest et de l'audit pour se prémunir face au risque cyber ?

Aujourd'hui, les grandes entreprises ont pleinement pris conscience du risque cyber. La sécurité fait partie intégrante des projets qui sont déployés en interne et la réalisation d'un audit de sécurité ou d'un pentest fait partie de la roadmap du projet. En 2008, lorsque j'ai démarré ma carrière de pentester, il y avait encore un important travail d'évangélisation à réaliser afin d'expliquer la différence entre un pentest et un audit notamment. Quinze ans plus tard, nous entrons directement dans le vif du sujet, à savoir quelle brique de l'infrastructure le client souhaite tester, quelles sont ses principales craintes par rapport à chaque brique — une perte de données ou une indisponibilité de l'infrastructure ? Souhaite-t-il réaliser une revue de configuration des serveurs ou de l'architecture ?



Cette maturité s'observe-t-elle aussi du côté des PME ?

Les PME maîtrisent moins le sujet, mais elles progressent. Selon moi, le Règlement Général sur la Protection des Données (RGPD) a permis de sensibiliser les PME à la cybersécurité, au moins sur la partie des données à caractère personnel. Le fait que la CNIL puisse infliger des sanctions financières en cas de manquement force les entreprises de manière générale à prendre en compte cette problématique. Auparavant, lorsqu'un pentester démontrait une vulnérabilité permettant de compromettre des données à caractère personnel, la réaction des clients était souvent de dire : « *oui c'est gênant, mais finalement, ce n'est pas si grave.* » Depuis l'entrée en vigueur du RGPD (mai 2018) et les sanctions financières prévues en cas de manquement, plus personne n'adopte cette position.

Donc globalement, lorsque je parle d'injection SQL ou de Cross-Site Scripting, beaucoup de clients connaissent ou ont déjà entendu parler de ces vulnérabilités. Les intervenants que nous rencontrons sont plus sensibilisés à la cybersécurité. Il est important, quand vient le moment de décrypter l'attaque devant le client, de ne pas le noyer de détails techniques, pour ne pas le perdre ou l'induire en erreur sur la manière de corriger. Un bon pentester doit savoir vulgariser. ■

GÉREZ VOTRE IT DE BOUT EN BOUT AVEC LA PLATEFORME EASYVISTA !

- IT SERVICE MANAGEMENT
- IT MONITORING
- REMOTE SUPPORT
- IT SELF-SERVICE
- DISCOVERY AND DEPENDENCY MAPPING

Rapide à déployer, facile à configurer, notre plateforme offre l'un des meilleurs coût total de possession (TCO) du marché.





Choisissez la solution
ERP **la plus complète**
intégrant nativement
l'IA et l'ESG*

*Élu meilleur ERP Software 2023 par le Palmarès de l'Informaticien