



Threat Hunting
Report 2024 de
CrowdStrike
Résumé

CROWDSTRIKE

Points clés du pionnier du renseignement
sur les cyberadversaires

Obtenir le rapport ↗

Accueil > Data Sciences, Machine Learning, Deep Learning, LLM

ACTUALITES

ILLUIN Technology place nAIxt en chef d'orchestre l'IA et de la GenAI

La startup française spécialiste de l'IA mise sur une plateforme « low-code » agnostique pour contrôler les transformations de données inférées par des modèles d'IA et d'IA générative. Une solution qui n'aurait pas réellement d'équivalent sur le marché.



par Gaétan Raoul, LeMagIT

Publié le: 18 déc. 2024

Reconnue pour ses projets de recherche en intelligence artificielle, ILLUIN Technology a développé des services d'accompagnement des grands groupes, mais aussi une suite de produits. À commencer par Dialogue, un outil développé à partir de 2018 pour concevoir des agents conversationnels. Il a été déployé à grande échelle par de grands groupes français, dont Enedis, EDF, La Macif, Boursobank, Randstad ou encore Geopost. La société revendique une soixantaine de clients.

Depuis, ILLUIN a étoffé son offre avec différents services consacrés aux traitements intelligents des documents – Doc Automation, Mail Automation, Search – mais aussi d'analyse de la voix et de l'intention client à chaud et à froid – Voice Parser et Speech Analyser.

« Sur le segment du traitement intelligent de documents, nos grands clients nous ont soumis à des benchmarks assez solides », déclare Robert Vesoul, cofondateur et CEO d'ILLUIN Technology. « Nos solutions savent traiter les documents les plus complexes : une facture d'hospitalisation, un bulletin de salaire, un constat d'accident automobile, etc. ».

Les noms de ces produits ont évolué. Ils ont eux-mêmes largement été « remodelés » sous l'influence de l'IA générative, selon le dirigeant.

PUBLICITE



MODULYS XM

ASI flexible et ultra-fiable
jusqu'à 600 + 50 kVA/kW

» DÉCOUVRIR

Actualités

ModernBERT : Answer.ai et Ligt turbo à BERT

Une transformation digitale n'est IT : les entreprises l'auraient ou

Abandon de VMware : Enix à év situation

Voir tous Nouve

Il y a toutefois une constante depuis quatre ans. Ils reposent tous sur un « backbone » principal. « Nous avons décidé en 2023 d'en faire un produit à part entière », signale Robert Vesoul. « C'est le framework qu'on ouvre à l'ensemble de nos clients pour utiliser sur des cas de développement, de réalisation sur mesure, de projets d'IA et [d'IA Générative](#) ». Ce framework, c'est la plateforme nAixt. Elle est présentée comme un « studio low-code » devant couvrir la conception d'applications impliquant l'orchestration de modèles d'IA, génératifs ou non. La communauté scientifique parle de système d'IA composite.

Alors que Copilot et ChatGPT ont dominé la scène ces deux dernières années, certaines entreprises espèrent bénéficier de l'automatisation supplémentaire apportée par les LLM dans des processus métier. « J'étais récemment dans une intervention du sommet Positive AI. Il y avait des intervenants qui évoquaient le fait que les organisations les plus avancées en IA générative sont celles qui développent vraiment des applications sur mesure, cœur de métier », affirme Robert Vesoul.

Un manque d'outils spécifiques

Or l'outillage disponible pour ce faire demeure complexe ou limité. Des frameworks comme LangChain et LlamaIndex se sont imposés, mais le paysage d'outils s'agrandit et ces bibliothèques réclament l'intervention de développeurs ou de personnes familières avec Python [ou JavaScript](#). « Ces solutions open source évoluent rapidement, mais couvrent une partie du sujet, et leur développement a été peu inspiré en ce qui concerne l'intégration des systèmes d'information d'entreprise », estime Robert Vesoul. [LangChain](#) existe aussi dans une distribution commerciale plus complète en la matière que son projet open source.

Historiquement, la majorité des équipes analytiques et de data science comptent des membres qui ne sont pas formés à Python et à JavaScript. Ceux-là et les externes utilisent ou ont longtemps utilisé des outils dotés d'une interface WYSIWYG pour gérer des pipelines de données. Le fait de traiter les données comme du code demeure un concept relativement nouveau et la barrière à l'adoption des [pratiques DevOps/MLOps](#) reste élevée.

« L'orchestration d'un framework en Python dans une organisation qui regroupe des data scientists, des data engineers et des data architects constitue un défi de taille », illustre Robert Vesoul. « Peu d'entreprises parviennent à le faire de manière efficace ».

Les fournisseurs cloud tentent d'y remédier, mais certaines entreprises, dont une partie des clients d'ILLUIN, outre l'enfermement propriétaire, redoutent les conflits d'intégrations entre des solutions parfois concurrentes. « Certains de ces outils dans les différents clouds dépendent des partis pris, qui ne sont pas nécessairement ceux de l'organisation pour d'autres projets de développements », note Robert Vesoul.

À cela s'ajoute la nécessité de gérer plusieurs grands modèles de langage, mais aussi différents algorithmes en tout genre. « Il y a la question de la performance, mais aussi des coûts », ajoute-t-

Télécharger Information Sé



DANS CE NUMÉRO:

- Magazine Information Sécurité commencent les cyberattaques
- Chakib Gzenayi : « L'EASM et le renseignement est plus im
- Michelin face à la problématique de la surface d'attaque exposée

Télécharger cette

Lire également

FQuAD : la langue française a e
...
- LeMagIT

Cinq exemples pour appréhender un ...
- LeMagIT

La R&D d'EDF se branche sur l'
- LeMagIT

il. « Aujourd’hui, les entreprises en 2024-2025 ont listé (ou terminent de lister) les cas d’usage prioritaires. Elles doivent maintenant organiser, développer, orchestrer cela à l’échelle de l’organisation ». Et la première case à cocher n’est autre que celle de la consommation des tokens : le rapport performance-coût.

« La plupart des grands clients qu’on interroge nous disent, “un de nos enjeux, c’est de profiter de ces modèles, mais tout en restant indépendant” », relate le CEO d’ILLUIN Technology.

Agnosticité, le maître mot

C’est l’un des arguments phares de la startup pour convaincre les grands groupes. Sa plateforme doit orchestrer des flux de travail propulsés à l’IA « hybride » de manière agnostique, que ce soit sur les trois grandes infrastructures cloud (GCP, Azure, AWS), sur les clouds souverains et qualifiés SecNumCloud (dont OVHcloud), mais aussi sur site. « Il s’agit de prouver que nous pouvons opérer des cas d’usage sans le recours aux API d’OpenAI, déployer rapidement des capacités et aussi changer des composants technologiques issus de différents fournisseurs en cours de route d’un projet », précise Ghislain Jeanneau, directeur produit nAIxt chez ILLUIN Technology.

Techniquement, les clients et les partenaires d’ILLUIN s’y retrouvent. C’est le socle technique des produits évoqués plus haut. C’est aussi une solution familière pour les équipes Data & IA.

Selon la présentation de Ghislain Jeanneau, nAIxt structure un projet en deux phases. Il y a d’abord la configuration dans le studio, puis l’exécution, c’est-à-dire la transformation des données « en résultats exploitables » via des API déployées.

Les usagers manipulent des projets comprenant des objets réutilisables – des prompts, du vocabulaire, etc. « Des templates de projets permettent de commencer à travailler rapidement sur des cas d’usages fréquents, comme le RAG, les traitements documentaires, un paradigme orienté agents », etc. », assure Ghislain Jeanneau.

Le développement dans nAIxt s’articule autour de deux types d’objets. Il y a d’abord un objet de type pipeline permettant de décrire les traitements en glisser-déposer en quelques manipulations. L’objet « test sets » rassemble des données de tests sur lesquelles les pipelines vont être éprouvés à chaque modification. « À chaque changement effectué, le pipeline est de nouveau exécuté et permet de comprendre l’impact des changements effectués ».

Une fois le pipeline prêt, le bouton « déploiement » expose ce pipeline via une API. Une interface et des modules permettent d’évaluer les performances et l’explicabilité des résultats dans la phase de configuration et au cours de l’exécution. L’UI inclut une fonction d’estimation des coûts afin de déterminer à travers les phases itératives de tests de pipelines quels sont les modèles d’IA les plus pertinents à exploiter. En cas d’erreur, de problèmes ou pour les besoins de la RSSI, la plateforme conserve des logs d’audit.

Sur le papier, tout cela ressemble à un outil ETL/ELT (extraction, chargement et transformation de données). Mais ici, le T ne correspondrait pas uniquement au moteur de requêtes d’une base de données ou d’une technologie type Apache Spark, mais à différentes modalités de transformation, incarnée par l’inférence de modèles d’IA (LLM, VLM, [computer vision](#), NLP, etc.). Ces transformations peuvent avoir lieu à différents endroits d’un pipeline afin de structurer les résultats et peuvent être liées ensemble. Ils sont effectués par des opérateurs, des fonctions spécifiques qui appellent différents LLM et outils. ILLUIN dit disposer de plus de 150 opérateurs. Le code généré par la plateforme peut être exécuté à travers des fonctions en mode

« serverless » sur des services tiers de type [FaaS](#).

S'il s'agit éventuellement de créer un runtime, les entreprises peuvent accéder au code source des pipelines au besoin. L'outil de visual programming engendre arbitrairement du code TypeScript. « Nous avons choisi ce langage, car il peut gérer des charges de travail hautement asynchrones », explique le responsable produit.

nAIxt s'intègre à diverses plateformes, [dont Amazon Bedrock](#), Google Vertex AI, Azure AI, Microsoft Sharepoint, Google Drive, Salesforce, Dynamics 365, des bases de données vectorielles (Qdrant, Elasticsearch), des outils d'inférence de modèles (vLLM, TGI, Ollama), etc.

Les documents eux-mêmes sont manipulables dans le studio et un système d'explicabilité vise à lier les informations à leurs sources. Certains opérateurs appliquant la méthode K des plus proches voisins permettent par exemple de vérifier qu'un montant affiché dans une facture est proche de celui annoncé dans un contrat. Ici, ILLUIN recommande d'utiliser des modèles de langage-vision, une [architecture RAG](#) et sa technique ColPali, exploitant les modèles « openweight » Google PaliGemma pour créer des embeddings visuels permettant d'extraire plus fidèlement les données en provenance de PDF ou de JPG.

« Les cas d'usage d'IA générative se concentrent souvent sur des applications conversationnelles, où le résultat final est généralement une simple chaîne de caractères ou un JSON », note Ghislain Jeanneau. « Les outils permettant un prétraitement ou un post-traitement avancé des données selon des conditions spécifiques font défaut », assure-t-il. « Cela limite les possibilités à des automatisations de bout en bout, souvent orientées vers des workflows spécifiques. Avec l'IA générative, ces scénarios sont moins directs que des approches traditionnelles, ce qui rend les outils d'orchestration indispensables pour atteindre un niveau supérieur d'automatisation ».

ILLUIN Technology croit avoir trouvé le moyen de se démarquer

Des acteurs français ou d'origine française disent cocher la cache agnostique, [notamment LightOn](#) et [Dataiku](#). « Selon ma compréhension, Lighton se positionne davantage sur des assistants, l'interaction utilisateur et le chaînage simple de tâches, tandis que Dataiku, issu d'un ADN plus traditionnel, reste axé sur l'entraînement de modèles, le MLOps et l'analyse classique » compare Robert Vesoul. « NAIxt cherche à unifier ces approches avec un studio de développement capable de gérer l'ensemble des cas d'usage de l'IA générative, des pipelines back-office aux outils conversationnels, là où les solutions existantes sont souvent limitées ». La plateforme d'ILLUIN serait alors complémentaire à ces deux solutions. Et plutôt en avance sur le marché.

Reste que des acteurs comme Microsoft, AWS, Google Cloud et Salesforce veulent ou peuvent traiter des processus de ce type. Pour l'heure, ils ont davantage musclé des assistants avec des outils et une architecture RAG pour en faire des sortes de « ShivaGPT ». Lors de la présentation officielle de nAIxt, le 4 décembre dernier, Geopost a présenté son intention d'utiliser l'IA générative pour extraire et analyser des événements liés aux colis à partir des logs du SI, afin de produire des résumés exploitables. Ces résumés pourront être présentés à des conseillers pour répondre aux clients ou directement intégrés dans des interactions via bots conversationnels, sans forcément exploiter une architecture RAG. « Les organisations matures sur ces sujets cherchent désormais des outils adaptés pour mettre en œuvre ce type de processus de bout en bout », insiste le CEO.

Pour l'heure, ILLUIN continue de cibler les « grandes organisations publiques et privées », mais

espère rassembler une plus large communauté en proposant dès 2025 une distribution open source de nAIxt. Un moyen également de s'étendre sur le marché ETI.

➤ Pour approfondir sur Data Sciences, Machine Learning, Deep Learning, LLM

IA générative : Informatica tente de se rendre indispensable  Par: Eric Avidon	La R&D d'EDF se branche sur l'IA générative  Par: Christophe Auffray	Lakeflow : Databricks veut unifier sa gestion des pipelines de données  Par: Gaétan Raoul
--	--	---

La couche de gouvernance de DBT Labs convainc davantage que son assistant IA  Par: Eric Avidon
--



À Propos	Annonces	E-Handbooks	Tous droits 2007 - 2024
Charte d'éthique et de déontologie	Partenaires	Conseils IT	Confidentialité
Rencontrez les journalistes	Dossier De Presse	Opinions	Paramètres
Contacts	Agenda	Guides Essentiels	
Utilisation Des Cookies	Nos Journalistes et Experts	Projets IT	
Réimpressions	Technologies		



Ce site web est détenu et exploité par Informa TechTarget, qui fait partie d'un réseau mondial qui vise à informer, orienter et mettre en relation les acheteurs et les vendeurs de technologies du monde entier. Tous les droits d'auteur leur appartiennent. Le siège social d'Informa PLC est situé au 5 Howick Place, Londres SW1P 1WG. Enregistré en Angleterre et au Pays de Galles. Le siège social de TechTarget, Inc. est situé au 275 Grove St. Newton, MA 02466.