

TECHNOLOGIE, VÉRITÉ ET CONFIANCE

Ouvrir la voie à une IA éthique

01 Synthèse	1
02 Pourquoi la réglementation de l'IA doit être « longuement et soigneusement examinée »	2
03 La vision de Nuix	2
04 Une IA qui fait ses preuves.....	3
05 L'IA et votre organisation.....	3
06 Résumé	4
09 Vous souhaitez en savoir plus ?	4

SYNTHÈSE

Ce livre blanc examine les moyens par lesquels les créateurs et les utilisateurs de solutions d'IA complexes peuvent réduire les nombreux risques éthiques posés aux organisations et à la société, sans compromettre la capacité de l'IA à contribuer à la résolution de nos défis les plus épineux et les plus urgents. Alors que l'évolution technologique s'accélère, il devient de plus en plus important de trouver un juste équilibre entre les promesses impressionnantes des progrès de l'IA et l'atténuation des menaces potentielles qu'ils font apparaître. En ce sens, l'IA illustre parfaitement l'expression « arme à double tranchant ».

Les deux tranchants de cette arme ont été aiguisés à la suite du lancement du ChatGPT par OpenAI en novembre 2022. Ce point d'inflexion important dans le développement de l'IA, qui s'étend sur plusieurs décennies, a permis de renforcer considérablement l'innovation, les investissements et la sensibilisation du grand public, tout en tirant davantage la sonnette d'alarme. Les préoccupations croissantes concernent de nombreux acteurs, tels que les entreprises, les salariés, les citoyens, les universités, l'administration publique et les autorités de réglementation. Étant donné la rapidité avec laquelle les choses évoluent, ces préoccupations sont compréhensibles. Peut-on affirmer que l'IA ne dépassera pas un jour notre capacité à la comprendre ou à la contrôler ?

Là encore, la perception que les gens ont de l'IA va dans les deux sens. Alors que 44 % des dirigeants interrogés par KPMG en 2021 estimaient que leur secteur avançait trop vite dans l'adoption de l'IA, la quasi-totalité des personnes interrogées souhaitaient que leur organisation avance encore plus vite.¹ En réalité, cette dichotomie s'est encore intensifiée.

Forte de plusieurs décennies d'expérience collective en matière d'IA et d'une compréhension approfondie des risques et des exigences de ses clients, Nuix reconnaît que le concept de confiance est au cœur du dilemme de l'IA. Encore récemment, la méfiance à l'égard de l'IA était l'apanage des films et des romans apocalyptiques. Nos contacts quotidiens avec l'IA se limitaient à acheminer nos appels téléphoniques, à corriger nos fautes d'orthographe ou à nous recommander des films. Mais alors que nous envisageons de confier à l'IA des tâches de plus en plus complexes et importantes, la crédibilité et l'exactitude de ses résultats font l'objet d'un examen de plus en plus minutieux, et ce, à juste titre.

« La plus grande source de sécurité pour l'homme, c'est la confiance ». Tel est le message incisif au cœur du remarquable discours de Jeffrey Bleich, directeur de Nuix, ancien ambassadeur des États-Unis en Australie et conseiller spécial du président Obama. Le discours de Bleich, prononcé devant des dirigeants d'entreprise en avril 2023, a souligné l'interconnexion importante entre la technologie, la vérité et la confiance dans notre société. L'IA ne pourra jouer un rôle aussi important dans la société que si elle est soumise à des normes éthiques strictes et si elle produit des résultats fiables et vérifiables.

Nuix se conforme à trois principes clés de conception technologique lors du développement de l'IA, qui correspondent à l'objectif de Nuix de construire une technologie qui soit une force au service du bien. Ces principes permettent de trouver le juste équilibre entre le respect des normes éthiques les plus strictes en matière d'IA dans nos solutions et l'optimisation de l'immense potentiel de l'IA pour nos clients.

Nous préconisons que ces trois principes soient au cœur de la conception et de l'utilisation de toute IA. La confiance de la communauté et de la société est ainsi placée au cœur de tous les progrès actuels et futurs. Ces trois principes sont les suivants :

Explicabilité – Pour obtenir et maintenir la confiance dans les résultats de l'IA, la « boîte noire » doit être ouverte pour donner à l'homme un aperçu clair des données d'entraînement utilisées pour créer les modèles, des biais qui pourraient exister et des raisons évidentes qui sous-tendent les prédictions de l'IA.

L'IA devrait être utilisée pour informer et accélérer la prise de décision humaine, et non pour la remplacer en fournissant des « réponses » définitives. L'IA de Nuix oriente les décideurs humains vers les endroits où ils sont le plus susceptibles de trouver les faits indéniables qui guideront leurs actions.

Accessibilité – Le pouvoir de transformation de toute technologie réside dans son utilisation et sa personnalisation. L'IA de Nuix permet à l'humain de garder le contrôle et de définir ce que la machine peut ou ne peut pas voir et faire en son nom, sans être obligé de recourir à des algorithmes ou à des codes inaccessibles. L'IA qui accroît les capacités humaines, et non qui les remplace, est au cœur de ce principe.

Spécificité – Le concept de spécificité favorise l'utilisation pragmatique et efficace de l'IA, qui permet d'obtenir des résultats mesurables dans le monde réel. Les modèles plus généraux sont plus difficiles à expliquer et moins transparents, ce qui les rend également moins reproductibles. Il est essentiel de permettre aux utilisateurs de piloter le processus en intégrant leurs connaissances et leur expertise dans les modèles créés et centrés sur l'un des nombreux cas d'utilisation ciblés. Il en résulte une amélioration considérable de la précision et une plus grande crédibilité des résultats propres au domaine.

Les effets de ces principes et la manière dont Nuix répond à ces exigences sont examinés plus loin dans le présent livre blanc.

POURQUOI LA RÉGLEMENTATION DE L'IA DOIT ÊTRE « LONGUEMENT ET SOIGNEUSEMENT EXAMINÉE »

Malgré la baisse du coût du développement et de la formation à l'IA en quelques années seulement, les ressources informatiques et humaines nécessaires restent importantes. Les progrès considérables en matière d'IA se sont donc concentrés dans les entreprises privées. En réalité, le monde universitaire ne produit qu'un dixième des percées² en matière d'IA. La plupart des applications et du développement de l'IA sont donc davantage guidés par la recherche du profit que par des préoccupations d'ordre éthique.

Les principaux innovateurs et défenseurs de l'IA appellent à une approche plus mesurée et plus prudente, ce qui prouve que la technologie évolue plus rapidement que la société n'en comprend les avantages et les risques. Pourtant, alors que des groupes de chefs d'entreprise proposent publiquement et collectivement un moratoire sur le développement de l'IA à une fréquence alarmante³, personne ne veut être le premier à reculer.

L'administration publique et les autorités de réglementation s'efforcent de suivre le rythme du changement et envisagent l'IA dans sa globalité, plutôt que de se contenter d'une réflexion superficielle. L'Union européenne, par exemple, est sur le point de promulguer l'une des premières lois globales sur l'IA. Le projet de loi sur l'IA⁴ a franchi une étape importante au Parlement européen quelques jours avant la publication de ce livre blanc. Cette loi pourrait être promulguée dès la fin de l'année 2023. Elle vise à « assurer un développement éthique et centré sur l'humain » de l'IA. Ce projet de loi repose sur la « capacité à comprendre les modèles utilisés pour obtenir les résultats, comment ces modèles ont été construits (données d'entraînement) et si le ou les humains de la boucle sont capables ou non d'être les arbitres finaux des résultats. »

Alors que l'UE a montré la voie en matière de protection des données grâce à l'adoption du RGPD⁵, il semble qu'elle soit à nouveau à l'avant-garde en ce qui concerne la réglementation de l'IA. Si l'on considère le passé comme un prologue, la loi sur l'IA de l'UE est un indicateur probable de la manière dont la réglementation de l'IA s'appliquera à l'échelle mondiale. En effet, elle fait déjà partie de la planification future de certaines organisations. Lorsqu'un client de Nuix – une organisation chargée de l'application de la loi en Europe – s'est renseigné sur notre IA, Nuix s'est fait un plaisir d'offrir des garanties écrites détaillées selon lesquelles ses modèles et capacités d'IA respectaient déjà les exigences.

Dans le cadre d'une approche plus ciblée visant à réduire les risques de l'IA, un juge fédéral du Texas a récemment adopté une nouvelle règle qui exige que les avocats comparaissant devant le tribunal soumettent un certificat attestant qu'aucune partie de leur dossier n'a été générée par l'IA générative (ChatGPT, Harvey.AI, Google Bard), ou que tout contenu assisté par l'IA a été minutieusement examiné par un humain pour s'assurer de son exactitude.⁶

En résumé, la réalisation du vaste potentiel de l'IA nécessitera à la fois une réflexion sérieuse et des actions réfléchies et coordonnées. Deux universitaires australiens spécialisés dans l'IA ont déclaré : « Il est clair que nous nous trouvons à un point d'inflexion : nous devons réfléchir sérieusement et de toute urgence aux inconvénients et aux risques de l'utilisation croissante de l'IA⁷. »

« L'aspect le plus triste de la vie actuelle est que la science acquiert des connaissances plus rapidement que la société n'acquiert de la sagesse. »

Isaac Asimov

LA VISION DE NUIX

Chez Nuix, nous utilisons et développons l'IA depuis des années. Nous estimons que l'IA peut transformer nos clients et réorganiser des économies entières. Cependant, l'humanité n'a peut-être qu'une seule chance de réussir la transition vers un avenir basé sur l'IA. Une réglementation intelligente, l'éducation, la discipline et un débat honnête sont essentiels pour garantir que l'IA est développée et utilisée de manière réfléchie, éthique et transparente. En résumé, l'IA éthique est synonyme d'IA de confiance.

Comme le souligne Stephen Stewart, directeur technique de Nuix : « En réalité, il est difficile de tirer parti de l'IA/ML dans l'incroyable diversité des défis liés aux données. Il n'existe pas vraiment de solution unique pour les données non structurées. L'IA de Nuix est un exemple d'application de concepts de pointe et de progrès technologiques pour répondre aux besoins très spécifiques des clients avec une efficacité incroyable et des informations fiables. »

Notre approche du développement et du déploiement de l'IA de confiance est régie par trois principes de conception technologique : **Explicabilité, Accessibilité et Spécificité**. Ces principes nous permettent de poursuivre notre objectif d'être une force au service du bien en trouvant la vérité dans les données numériques. Ils veillent à ce que nous créions une intelligence collaborative qui accroît l'efficacité et l'efficacité humaine, permettant aux personnes de se concentrer sur les tâches à forte valeur ajoutée qui requièrent un jugement humain, une pensée critique nuancée et une conscience éthique.

Ces principes font de Nuix un fournisseur d'IA responsable, fiable et éthique. Ils servent également de lignes directrices pour guider nos efforts de développement continue et garantir que nous restons en phase avec les exigences des clients et l'évolution des normes éthiques mondiales en matière d'IA.

1. EXPLICABILITÉ

En tant que leaders mondiaux des logiciels d'intelligence numérique, nos clients font appel à nos services pour traiter, passer au crible et interpréter de vastes quantités de données afin de trouver des informations factuelles, c'est-à-dire la vérité. Pour instaurer et maintenir cette confiance dans l'IA, nos clients doivent avoir accès aux données sous-jacentes utilisées pour construire nos modèles d'IA, ainsi qu'aux explications qui sous-tendent les résultats produits par ceux-ci. C'est ce que l'on appelle de plus en plus souvent dans le secteur l'« explicabilité », et notre technologie d'IA a été conçue dès le départ à cette fin.

Un autre élément important de l'explicabilité est le fait que notre IA ne cherche pas à prendre de décisions à la place de l'homme. Elle fournit plutôt des données triées, pertinentes et hiérarchisées qui soutiennent une prise de décision humaine plus efficace et efficiente⁸ de la part de nos clients. Il est essentiel que l'homme dispose d'informations directes à chaque étape du processus et qu'il garde le contrôle total en tant qu'arbitre final. Notre IA offre quatre couches d'analyse principales pour les données textuelles :

1. Type d'enregistrement : document fiscal, formulaire de soins de santé, relevé de notes, brevet, etc.
2. Sujet du contenu : droit, environnement, finance, politique, etc.
3. Extractions : PII, PHI, entités nommées, etc.
4. Priorité : pertinence, importance ou niveau de risque de l'un des éléments ci-dessus, sur la base de règles définies par le client, etc.

La solution Nuix oriente les décideurs humains vers les endroits où ils sont le plus susceptibles de trouver les faits indéniables qui guideront leurs actions et augmenteront la précision et l'efficacité de leur prise de décision. Chaque étape du processus est enregistrée, répétable, explicable et défendable.

2. ACCESSIBILITÉ

Encore récemment, l'IA était l'apanage d'un petit nombre d'experts en apprentissage automatique, de programmeurs de logiciels et d'experts en sciences des données hautement qualifiés. Ces circonstances ont exclu les experts en la matière (EM) de la chaîne de valeur de l'élaboration des modèles et les ont laissés trop dépendants des équipes techniques qui n'avaient souvent pas les connaissances nécessaires pour les utiliser.

Pour maximiser la valeur de l'investissement dans l'IA et répondre à la nécessité éthique d'étendre les opportunités et les avantages offerts par l'IA, il est nécessaire de permettre aux personnes non spécialisées d'accéder facilement et intuitivement aux fonctionnalités de l'IA.

Nuix contribue à révolutionner l'expérience utilisateur grâce à une interface intuitive conçue pour permettre aux experts du domaine de construire, d'optimiser et de valider rapidement des modèles d'IA linguistique très précis – une interface pointer-cliquer simple et sans une seule ligne de code. Au-delà des progrès en termes de flux de travail et de productivité, ce niveau d'accessibilité renforce la confiance dans le fait que les résultats de l'IA sont vérifiés et bien alignés sur le sujet pour lequel ils ont été conçus.

Des interfaces bien conçues avec peu ou pas de code et des flux de travail intégrés favorisent la démocratisation de l'IA, en permettant à un plus grand nombre de travailleurs humains d'interagir directement avec l'IA pour apporter les modifications nécessaires, tester et optimiser les modèles, corriger les erreurs et supprimer les biais. Ainsi, c'est l'utilisateur de l'IA et non le programmeur qui garde le contrôle ultime.

3. SPÉCIFICITÉ

Dans le domaine des applications réelles, plus grand n'est pas toujours synonyme de meilleur. Malgré le soutien de ChatGPT par GPT4 (le plus grand LLM jamais construit⁹), son immensité et sa volonté de plaire à tout le monde compromettent une grande partie de son utilisation pratique dans des environnements commerciaux réels. Il est évident qu'il serait peu pratique, voire impossible, d'héberger ChatGPT dans votre environnement sécurisé. De plus, le téléchargement de données sensibles et leur utilisation « en l'état » sur le Web présentent un risque d'exposition important. Pour cette seule raison, de nombreuses organisations ont interdit à leurs employés d'accéder à ChatGPT et à d'autres grands sites de fournisseurs d'IA générative. Ces modèles de langage massifs ne sont pas bien adaptés aux sujets spécifiques et pertinents pour votre entreprise ou votre domaine d'activité.

En revanche, l'IA de Nuix est conçue pour être déployée dans les environnements sécurisés des clients et fonctionner sur des systèmes abordables basés sur des processeurs. De plus, si l'on combine les avantages de l'explicabilité et de l'accessibilité décrits ci-dessus, les clients peuvent rapidement adapter les modèles préétablis, voire en créer de nouveaux, pour répondre à leurs exigences subjectives et à leurs cas d'utilisation spécifiques. Là encore, ce sont les humains qui ont le contrôle, et l'IA s'efforce de répondre à leurs besoins spécifiques.

« Un ordinateur mériterait d'être qualifié d'intelligent s'il pouvait tromper un humain en lui faisant croire qu'il était humain. »

Alan Turing

UNE IA QUI FAIT SES PREUVES

Ces trois principes de conception technologique nous ont permis de remporter le « CUI Tool Automation Challenge » de l'US Navy, annoncé en décembre 2022¹⁰.

L'objectif de ce concours était de soutenir le partage approprié, efficace et essentiel d'informations sensibles mais non classifiées – Controlled Unclassified Information (CUI) – entre les services, les partenaires et les alliés. Les tâches spécifiques à gérer étaient les suivantes :

- l'ingestion de millions de documents du gouvernement américain et du ministère de la Défense dans différents formats de fichiers
- l'analyse du contenu des documents afin d'identifier les types de CUI dans chaque fichier
- la quantification d'un niveau de confiance en fonction de chaque résultat
- le marquage du bandeau et du pied de page de chaque document afin de maintenir la chaîne de confiance lors de l'interaction entre les différents utilisateurs et les fichiers.

Le succès de notre solution repose sur les fondements de nos trois principes. L'**explicabilité** fournie par une interface simple de type pointer-cliquer sur la façon dont les modèles d'IA existants et récents ont été créés, ainsi que les raisons facilement compréhensibles à l'origine de nos résultats. L'**accessibilité** offerte par une interface sans code, qui permet à l'utilisateur de construire, d'entraîner, d'améliorer ou d'adapter les modèles et les pondérations relatives afin d'optimiser et de valider leurs performances. Enfin, la **spécificité** de chaque modèle, soigneusement mis au point pour un type spécifique de CUI, et combiné par des experts afin de relever des défis uniques en matière de données réelles.

L'IA ET VOTRE ORGANISATION

Selon certains économistes, la capacité à utiliser rapidement et efficacement les technologies de pointe est plus déterminante pour la croissance que la nature même de la technologie. La célèbre phrase de Robert Solow est très pertinente : « On voit des ordinateurs partout, sauf dans les statistiques de productivité ».

Chez Nuix, nous nous sommes forgés une réputation mondiale en rendant les données de nos clients hautement consultables. Aujourd'hui, nous passons à la vitesse supérieure en proposant une plateforme unifiée et imprégnée d'IA appelée Nuix Neo. Nuix Neo intègre le moteur de traitement de données le plus puissant au monde et l'IA avancée pour résoudre les défis complexes et croissants de nos clients en matière de données. Grâce à cette stratégie d'innovation intégrée, les clients de Neo peuvent tirer parti de l'IA, sans avoir besoin d'une expertise interne en matière d'IA ni d'acheter ou de construire des solutions d'IA ponctuelles.

RÉSUMÉ

Proposer des solutions d'IA est une tâche qui ne doit pas être prise à la légère. Dans un monde en proie à des Chatbots hallucinants et bavards et assailli par la désinformation et les fausses informations, les enjeux sont plus importants que jamais. Toutefois, les problèmes d'aujourd'hui et de demain ne peuvent tout simplement pas être résolus avec les techniques et les technologies d'hier. Nous estimons que la maîtrise de l'IA est indispensable pour assurer un avenir prospère et durable.

Le clé sera une approche réfléchie et disciplinée de l'IA, fondée sur l'explicabilité, l'accessibilité et la spécificité. L'objectif ultime sera d'obtenir des résultats fiables, vérifiables et justifiables pour nos clients et la société en général.

VOUS SOUHAITEZ EN SAVOIR PLUS ?

Les tendances abordées dans ce livre blanc continuent d'évoluer, tout comme les applications concrètes dont disposent nos clients pour les fonctionnalités d'IA de Nuix. Pour en savoir plus sur la façon dont ces fonctionnalités peuvent être utilisées dans votre organisation, contactez-nous via les canaux décrits ci-dessous. L'un de nos experts en IA vous contactera pour vous aider à comprendre comment la solution Nuix Neo et son IA intégrée peuvent être utilisées dans votre organisation.

RÉFÉRENCES

- ¹ Krishna, Campana et Chandrasekaran, 'Thriving in an AI World: Unlocking the value of AI across seven key industries,' KPMG, 2021, <https://advisory.kpmg.us/articles/2021/thriving-in-an-ai-world.html>
- ² Stanford University, 'Chapter 1.2., Research and Development', Fig 1.2.2, Artificial Intelligence Index Report 2023', 2023, <https://aiindex.stanford.edu/report/#individual-chapters>
- ³ Voir <https://www.science.org/content/article/alarmed-tech-leaders-call-ai-research-pause#:~:text=Attracting%20signatures%20from%20the%20likes,potential%20risks%20of%20the%20technology>
- ⁴ Voir <https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence>
- ⁵ Voir <https://gdpr.eu/what-is-gdpr/>
- ⁶ Voir <https://www.mitchellwilliamsllaw.com/taming-the-ai-beast-judges-set-rules-to-control-use-of-generative-ai-in-their-courts>
- ⁷ Sonenberg et Walsh, 'Artificial intelligence is now part of our everyday lives and its growing power is a double edged sword', The Conversation, novembre 2021, <https://theconversation.com/artificial-intelligence-is-now-part-of-our-everyday-lives-and-its-growing-power-is-a-double-edged-sword-169449>
- ⁸ À l'heure où nous rédigeons ce livre, Nuix est en discussion avec l'un des principaux organismes européens chargés de l'application de la loi, qui cherche à savoir comment tirer parti des avantages de l'IA, tout en se concentrant sur la transparence des résultats.
- ⁹ Voir <https://neuroflash.com/blog/gpt-4-parameters-rumors-and-forecasts/#:~:text=It%20is%20also%20possible%20that,trillion%20or%2010%20trillion%20parameters>
- ¹⁰ 'Nuix and Serco NA partner to score prizewinning results in US Navy AI Automation Challenge', Nuix, <https://www.nuix.com/news/nuix-and-serco-na-partner-score-prizewinning-results-us-navy-ai-automation-challenge>



Nuix (www.nuix.com, ASX:NXL) crée des logiciels innovants qui permettent aux organisations de trouver simplement et rapidement la vérité à partir de toutes les données dans un monde numérique. Nous sommes une équipe passionnée et talentueuse, qui enchante nos clients grâce à des logiciels qui transforment les données en informations exploitables et les aident à relever les défis liés aux litiges, aux enquêtes, à la gouvernance, aux risques et à la conformité.

APAC

Australie : +61 2 8320 9444

EMEA

Royaume-Uni : +44 203 934 1600

AMÉRIQUE DU NORD

États-Unis : +1 877 470 6849

Nuix (et toutes les autres marques de Nuix utilisées) sont des marques de Nuix Ltd. et/ou de ses filiales, le cas échéant. Tous les autres noms de marques ou de produits sont des marques déposées de leurs détenteurs respectifs. Toute utilisation des marques de Nuix nécessite l'approbation écrite préalable du service juridique de Nuix. Vous pouvez contacter le service juridique de Nuix par e-mail à l'adresse Legal@nuix.com.

CE MATÉRIEL CONSTITUE LA PROPRIÉTÉ INTELLECTUELLE DE NUIX LTD. ET DE SES FILIALES (« NUIX »), Y COMPRIS LES ÉLÉMENTS PROTÉGÉS PAR LES DROITS D'AUTEUR QUI ONT ÉTÉ SIGNALÉS COMME TELS ET/OU ENREGISTRÉS AUPRÈS DU BUREAU DU DROIT D'AUTEUR DES ÉTATS-UNIS. TOUTE REPRODUCTION, DISTRIBUTION, TRANSMISSION, ADAPTATION, PRÉSENTATION PUBLIQUE OU REPRÉSENTATION PUBLIQUE DE LA PROPRIÉTÉ INTELLECTUELLE (AUTRE QU'À DES FINS INTERNES PRÉ-APPROUVÉES) NÉCESSITE L'APPROBATION ÉCRITE PRÉALABLE DE NUIX.