LIVRET D'ENTREPRISE

ALIA SANTÉ

PIONNIERS DE LA GÉNÉRATION DE DONNÉES SYNTHÉTIQUES EN EUROPE

Brief Introduction

Les données synthétiques représentent une avancée majeure pour le secteur de la santé.

Elles offrent une alternative sûre, éthique et performante à l'utilisation directe des données réelles, souvent sensibles et difficilement partageables.

Ce livre blanc vise à présenter clairement le fonctionnement, les avantages et les cas d'usage concrets de cette technologie, afin de vous accompagner en toute confiance dans son adoption.

Qu'est-ce qu'une donnée synthétique ?

Une donnée synthétique est une donnée générée à partir de modèles d'intelligence artificielle entraînés sur des données réelles.

Ces modèles apprennent les distributions statistiques sous-jacentes sans reproduire les individus.

Le résultat est une donnée d'apparence réaliste mais totalement dénuée d'informations personnelles.

Entrainez votre modèle

Procédez à la génération de données synthétiques

Obtenez une certification pour vos données

Les 3 grands bénéfices des données synthétiques

1. Protection des données personnelles
2. Enrichissement statistique des jeux de données
3. Facilitation du partage et de l'innovation

BÉNÉFICE 1

Protection des données personnelles

Enjeux

RGPD, secret médical, accès restreint aux données sensibles, exigences éthiques et réglementaires fortes dans le domaine de la santé.

Favorisez le développement de nouvelles solutions dans le respect de la règlementation grâce aux données synthétiques

BÉNÉFICE 1 - SUITE

Détail du bénéfice

L'accès aux données de santé réelles est soumis à des contraintes strictes, en particulier en Europe avec le RGPD et les recommandations des autorités sanitaires comme la CNIL. La moindre possibilité de ré-identification d'un individu, même indirecte, entraîne de lourdes restrictions.
Les données synthétiques offrent une alternative conforme et fiable, car elles ne contiennent aucune information personnelle issue directement d'un patient.

Leur production repose sur des algorithmes avancés d'intelligence artificielle capables de prendre en compte la statistique globale et non sur un simple masquage des données.
Ce caractère fondamentalement non ré-identifiable est vérifié à travers des métriques de confidentialité permettant d'évaluer la distance entre les données réelles et synthétiques calculées dans notre rapport de qualité.
Ainsi, les jeux de données synthétiques peuvent être partagés, stockés ou utilisés pour entraîner des modèles d'IA sans risquer d'exposer des données personnelles.

BÉNÉFICE 1 - SUITE

Anonymisation native conforme au RGPD

Les données synthétiques ne nécessitent pas de traitements supplémentaires pour garantir la conformité, car leur nature artificielle les rend non identifiables par construction.

Absence de lien direct ou indirect avec une personne physique

Chaque donnée générée est créée ex nihilo par un modèle IA et ne provient jamais d'un dossier patient spécifique.

Réduction du nombre d'autorisations nécessaires

Contrairement aux données réelles, les données synthétiques ne nécessitent ni autorisation CNIL, ni comité d'éthique pour être exploitées dans la plupart des cas d'usage, simplifiant ainsi les démarches administratives.

BÉNÉFICE 1 - SUITE

Cas d'usage

FORMATION MÉDICALE SANS CONTRAINTE RÉGLEMENTAIRE

Un CHU souhaite mettre à disposition des cas cliniques pour former ses internes. Grâce aux données synthétiques, il peut générer des situations réalistes, variées, et sans risque pour la confidentialité des patients. Cela permet d'enrichir les enseignements tout en respectant les normes éthiques.

SIMULATION RÉGLEMENTAIRE DANS UN LABORATOIRE PHARMACEUTIQUE

Une entreprise pharmaceutique souhaite entraîner un algorithme d'analyse de données cliniques avant le lancement d'un essai multicentrique. L'usage de données synthétiques issues d'études précédentes permet de tester différents scénarios sans manipuler de données sensibles, tout en préparant le pipeline analytique.

BÉNÉFICE 1 - SUITE

ACCÉLÉRATION DE LA R&D DANS UNE HEALTHTECH

Une startup développant une solution d'aide au diagnostic dispose d'un jeu de données limité et fortement déséquilibré. Grâce à la génération de données synthétiques, elle parvient à constituer un corpus d'entraînement robuste et conforme, tout en répondant aux exigences des investisseurs sur les délais et les coûts de développement.

CONFIDENTIALITÉ

Favorisez le développement de nouvelles solutions dans le respect de la règlementation grâce aux données synthétiques

ACCÉLÉRATION

Accélérez vos projets de recherche et développement et développez des solutions innovantes tout en réduisant les risques liés aux donnés réelles

PARTAGE

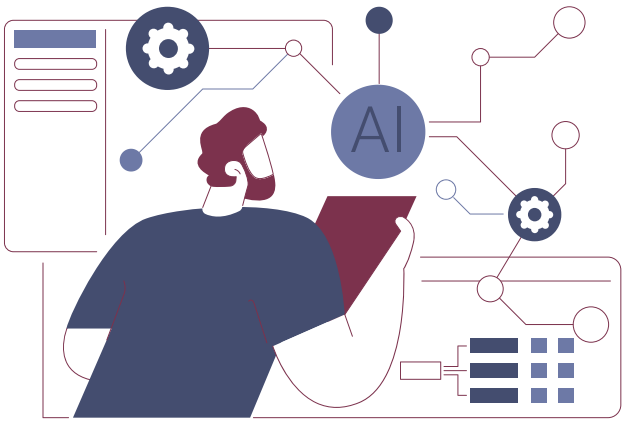
Facilitez le partage en éliminant les risques de divulgation d'informations sensibles et collaborez plus facilement

BÉNÉFICE 2

Enrichissement statistique des jeux de données

Enjeux

biais, données rares, déséquilibres de classes, sous-représentation de certaines populations ou pathologies.



Les données réelles disponibles sont souvent déséquilibrées

BÉNÉFICE 2 - SUITE

Détail du bénéfice

Certaines pathologies rares, profils cliniques spécifiques ou sous-groupes démographiques sont insuffisamment représentés.

Cela limite la capacité des modèles d'intelligence artificielle à généraliser correctement, augmentant les risques de biais ou d'erreurs de diagnostic.
La génération de données synthétiques permet de créer artificiellement des exemples statistiquement cohérents pour combler ces lacunes.

Ces données peuvent être calibrées pour représenter des cas minoritaires, simuler des scénarios critiques ou générer des données correspondant à des profils cliniques précis. En conséquence, les modèles d'IA gagnent en robustesse, en équité et en performance globale.

BÉNÉFICE 2 - SUITE

Génération de cas réalistes pour les classes sous-représentées

Les données synthétiques permettent d'équilibrer les datasets en créant des exemples issus de profils rares mais pertinents cliniquement.

Amélioration de la robustesse des modèles IA

Une meilleure représentativité statistique réduit le risque de sur-apprentissage ou de généralisation biaisée, surtout dans les modèles cliniques sensibles.

Possibilité de créer des scénarios extrêmes ou simulés

Permet d'explorer des situations rares ou critiques pour tester la fiabilité d'un algorithme ou simuler des réactions dans un essai virtuel.

BÉNÉFICE 2 - SUITE

Cas d'usage

DÉVELOPPEMENT D'UN MODÈLE IA POUR MALADIE RARE (MEDTECH)

Une startup développant une solution d'aide au diagnostic dispose d'un jeu de données limité et fortement déséquilibré. Grâce à la génération de données synthétiques, elle parvient à constituer un corpus d'entraînement robuste et conforme, tout en répondant aux exigences des investisseurs sur les délais et les coûts de développement.

ÉQUILIBRAGE DE DONNÉES CLINIQUES DANS UN LABORATOIRE PHARMACEUTIQUE

Un laboratoire conçoit un outil d'aide à la décision basé sur des dossiers patients hospitaliers. Certains profils (personnes âgées polypathologiques) sont sous-représentés. Les données synthétiques comblent ces manques pour tester la sensibilité du modèle à ces cas critiques.

BÉNÉFICE 2 - SUITE

SIMULATION D'ÉCHEC THÉRAPEUTIQUE POUR L'ÉVALUATION D'UN DISPOSITIF

Une Healthtech souhaite simuler des cas d'inefficacité d'un traitement afin de valider les capacités prédictives d'un dispositif médical. Elle utilise des données synthétiques représentant des profils de non-répondeurs cliniquement réalistes.

Or, les données disponibles sont trop limitées pour couvrir l'ensemble des cas cliniques nécessaires à l'entraînement du modèle. Grâce à la génération de données synthétiques ciblées, l'entreprise peut compléter son jeu de données initial, introduire davantage de diversité statistique et ainsi entraîner un modèle plus robuste et performant, validable dans un cadre réglementaire strict.

BÉNÉFICE 3

Facilitation du partage et de l'innovation

Enjeux

Cloisonnement des données, lourdeurs juridiques, frein à l'innovation collaborative, silos entre institutions, barrières à la mutualisation des données entre acteurs publics et privés.

Les données synthétiques permettent de contourner ces obstacles

BÉNÉFICE 3 - SUITE

Détail du bénéfice

Dans le secteur de la santé, l'exploitation des données est souvent entravée par des réglementations strictes, une gouvernance complexe et des barrières à la circulation de l'information entre les acteurs. La collaboration entre établissements de soins, laboratoires pharmaceutiques, centres de recherche ou startups est ainsi fortement limitée par les contraintes de confidentialité.

En remplaçant les données sensibles par des équivalents non identifiables mais statistiquement fidèles, elles autorisent un partage sécurisé des jeux de données, sans risque pour les patients ni pour la conformité réglementaire. Cela ouvre la voie à des dynamiques de co-innovation, à l'évaluation croisée de solutions d'IA, ou encore à la mutualisation d'expertises autour de bases de données communes

BÉNÉFICE 3 - SUITE

Déverrouille le partage inter-établissements ou public-privé

Permet de construire des bases de données accessibles à plusieurs partenaires sans compromettre la confidentialité.

Favorise l'ouverture de jeux de données pour la recherche

Accélère les collaborations scientifiques tout en assurant la sécurité juridique des échanges

Accélère la validation d'algorithmes

Donne accès à des environnements d'entraînement et de test réalistes, même lorsque l'accès aux données réelles est restreint.

BÉNÉFICE 3 - SUITE

Cas d'usage

CRÉATION D'UNE BASE DE DONNÉES INTER-ÉTABLISSEMENTS

Un consortium réunissant un CHU, une startup en IA santé et un laboratoire pharmaceutique souhaite entraîner un modèle prédictif sur une base de données partagée. Grâce aux données synthétiques générées à partir des jeux respectifs de chaque acteur, ils peuvent construire une base mutualisée conforme au RGPD et adaptée à la validation croisée.

OUVERTURE DE DONNÉES POUR UN APPEL À PROJETS COLLABORATIF

Un centre de recherche public souhaite répondre à un appel à projets européen nécessitant le partage de données cliniques. Les données synthétiques lui permettent de respecter les exigences de confidentialité tout en montrant la faisabilité du projet.

BÉNÉFICE 3 - SUITE

VALIDATION EXTERNE D'UN ALGORITHME DÉVELOPPÉ PAR UNE HEALTHTECH

Une startup ayant conçu un modèle d'aide à la décision souhaite le tester auprès de plusieurs hôpitaux. Ceux-ci sont réticents à transmettre des données réelles.

En générant des données synthétiques à partir de leurs bases internes, ils peuvent participer aux tests sans exposer les dossiers patients. veut créer une base de données commune.

Les données synthétiques permettent un partage sans risquer de compromettre les sources.

Acceptation réglementaire

Nos solutions reposent sur une infrastructure technologique d'intelligence artificielle, développée spécifiquement pour répondre aux exigences du secteur de la santé. Cette technologie à été développée sur un programme de recherche depuis 2022, co-financé par l'union Européenne.

La génération de données synthétiques chez Alia Santé repose sur des modèles d'intelligence artificielle avancés, incluant les GANs (Generative Adversarial Networks), les VAEs (Variational Autoencoders) et les modèles de diffusion, permettant une reproduction fidèle des structures statistiques complexes.

Nous adaptons systématiquement nos architectures aux types de données traitées (tabulaires, images médicales, séries temporelles, signaux biologiques) et aux objectifs des projets (entraînement de modèles, simulation, anonymisation, etc.).

Acceptation réglementaire

Chaque jeu de données généré fait l'objet d'un processus rigoureux d'évaluation, basé sur trois piliers fondamentaux :

Score de fidélité statistique

Comparaison quantitative entre les distributions synthétiques et réelles pour garantir la similarité des patterns et corrélations clés.

Score de confidentialité

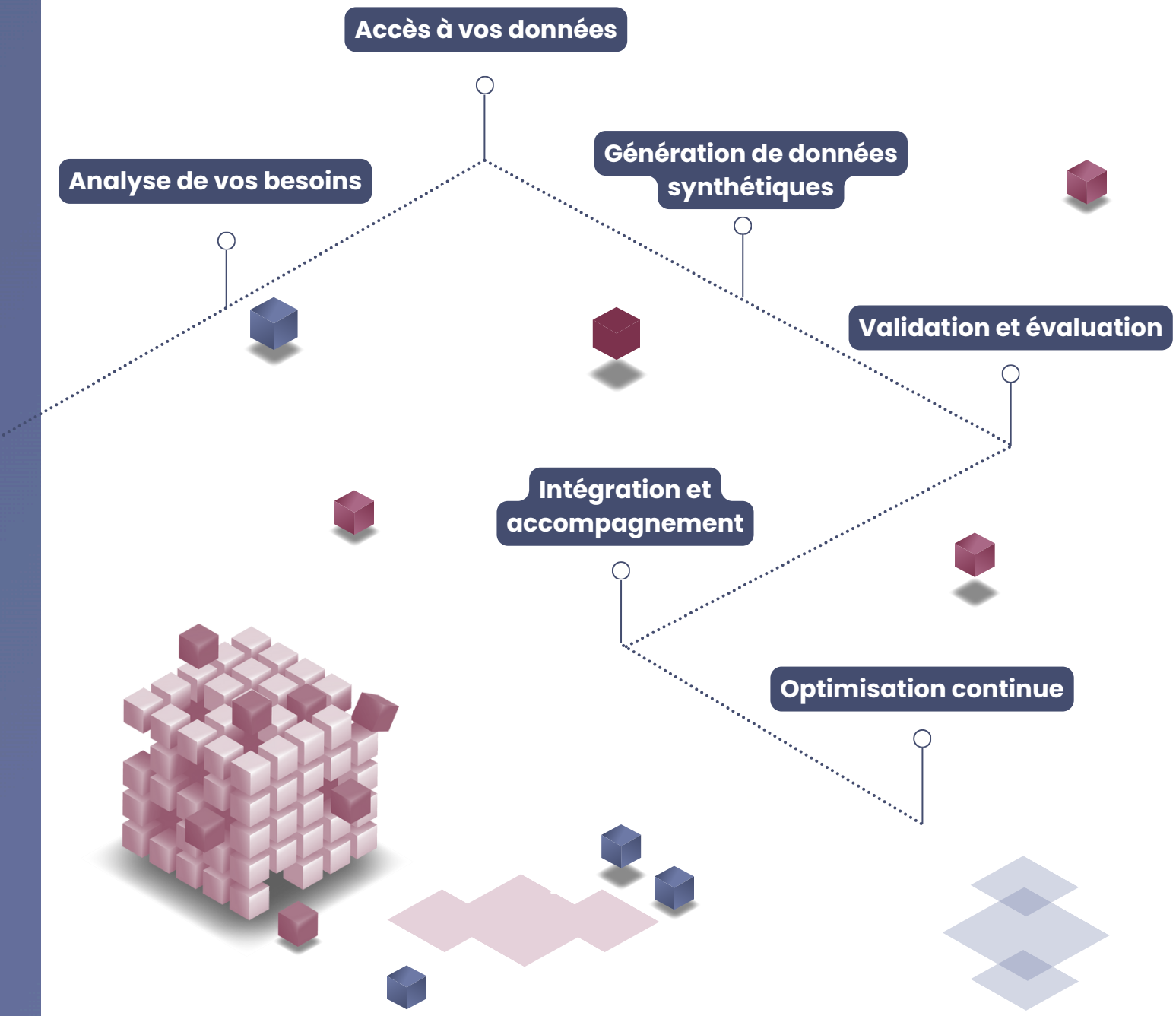
Tests de vulnérabilité et calculs de risques de ré-identification afin de démontrer l'innocuité des données générées.

Score d'utilité :

Nous développons également des modules de visualisation et d'interprétabilité pour permettre à nos clients de comprendre les comportements des modèles générateurs et d'ajuster les paramètres en toute transparence.

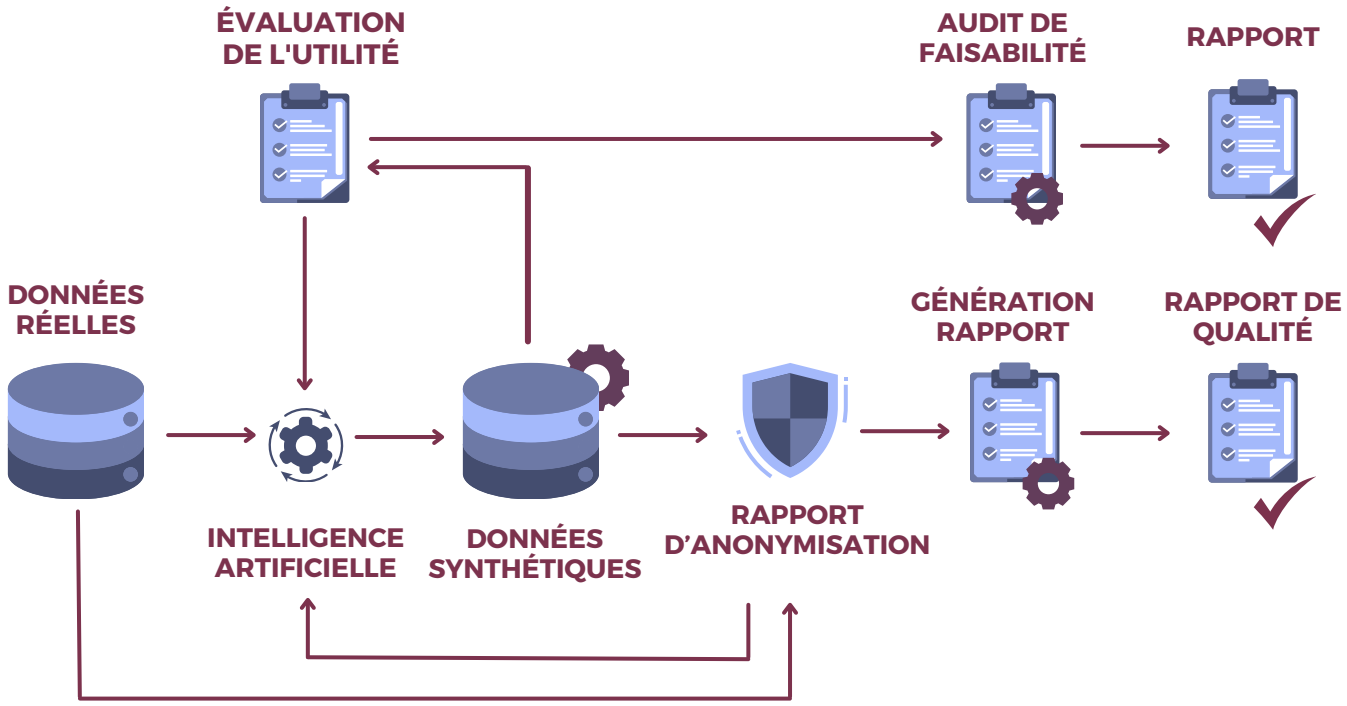
FONCTIONNEMENT PRODUIT

Méthodologie de génération de données



PROCESSUS DE GÉNÉRATION

Méthodologie de génération de données



Accompagnement personnalisé et sur mesure à la demande.

ÉTAPES ET LIVRABLES

Méthodologie de génération de données

Analyse de vos besoins

Nos experts travaillent avec vous pour comprendre vos objectifs spécifiques, vos contraintes et vos défis.

Accès à vos données

Transférez ou inter-connectez vos données en toute sécurité avec notre plateforme conforme RGPD & HDS.

Génération de données synthétiques

Lancez l'entraînement de vos modèles grâce à AliaDataGen pour produire des données fiables et anonymisées.

Validation et évaluation

Recevez un rapport détaillé sur la qualité et l'utilité des données générées pour garantir leur adéquation avec vos projets.

Intégration et accompagnement

Bénéficiez d'un support continu pour intégrer vos nouvelles données dans vos systèmes existants et maximiser leur valeur.

Optimisation continue

Suivez les performances et ajustez vos paramètres grâce à nos outils de visualisation et d'analyse avancée.

Pourquoi choisir Alia Santé ?

- **Une équipe scientifique composée de PhD, cliniciens et experts IA.**
- **Une expérience concrète dans la mise en œuvre de projets IA en santé.**
- **Une maîtrise des réglementations et des contraintes métiers.**
- **Une technologie opérationnelle, déjà déployée chez des clients.**

En collaborant avec Alia Santé, bénéficiez de notre agrément CIR (Crédit d'Impôt Recherche) et CII (Crédit d'Impôt Innovation).

-30% DE RÉDUCTION SUR LES DÉPENSES

Pourquoi choisir Alia Santé ?

Alia Santé est une startup toulousaine pionnière de la génération de données synthétique dont la raison d'être est de transformer le paysage de la santé grâce à l'utilisation éthique et innovante de l'intelligence artificielle.

+12 EXPERTS

Composent notre comité scientifique

+4 ANNÉES D'EXPÉRIENCES

En intelligence artificielle et data

+6 PRIX INNOVATIONS

Depuis 2 ans

+8 PARTENAIRES

Reconnus mondialement

TROPHÉES de l'économie NUMÉRIQUE 2023 · UNIVERSITÉ DE LA e-santé CASTRES-MAZAMET · La FRENCH CARE · NEXT INNOV · TECH FOR FUTURE

Conclusion

Les données synthétiques ne remplacent pas les données réelles : elles les complètent, les protègent et leur donnent une nouvelle valeur. Bien utilisées, elles sont un formidable levier d'accélération pour vos projets IA, dans le respect des principes éthiques et réglementaires.

Pour toute question ou pour démarrer une expérimentation avec nos données synthétiques, contactez-nous.

06 79 69 62 16

www.alia-sante.com

contact@alia-sante.com