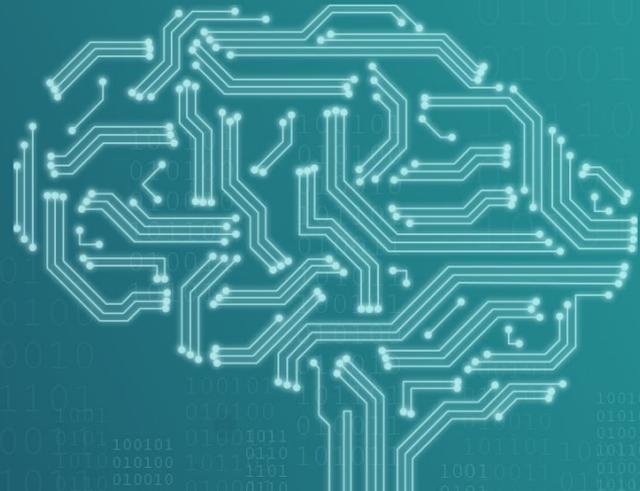


ATELIER 1

Comment l'IA améliore
le contrôle qualité
des filières industrielles
exigeantes ?



1.

Au travers de 12 cas d'usages : Comment l'IA améliore concrètement le contrôle qualité dans l'Industrie ?

2.

Comment est réalisé un projet d'IA, de l'idéation à l'industrialisation ?

3.

Les Données, la clé d'une IA efficace & l'Optimisation des modèles d'IA

4.

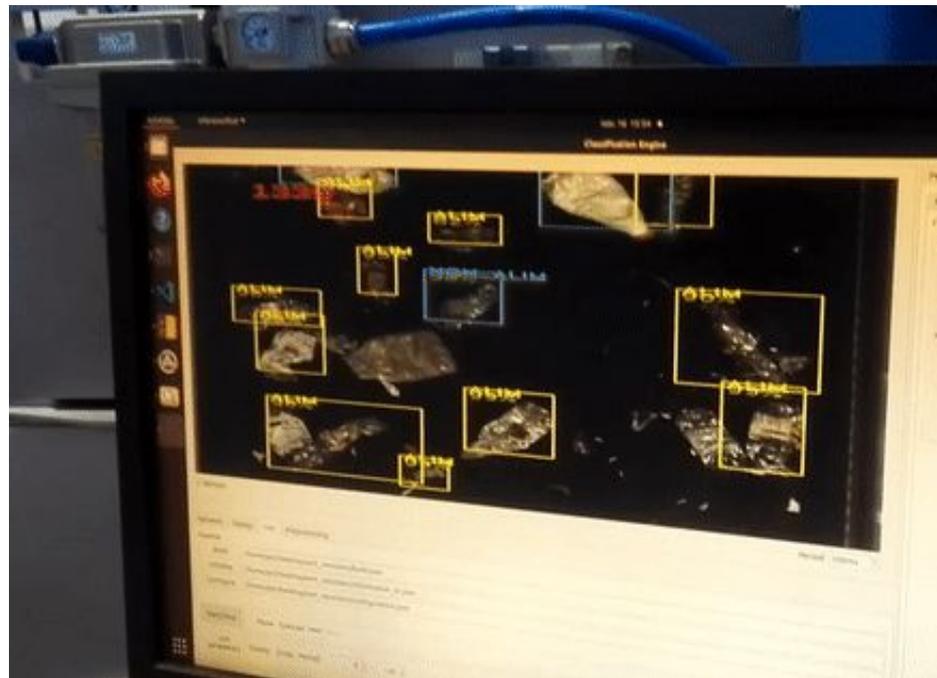
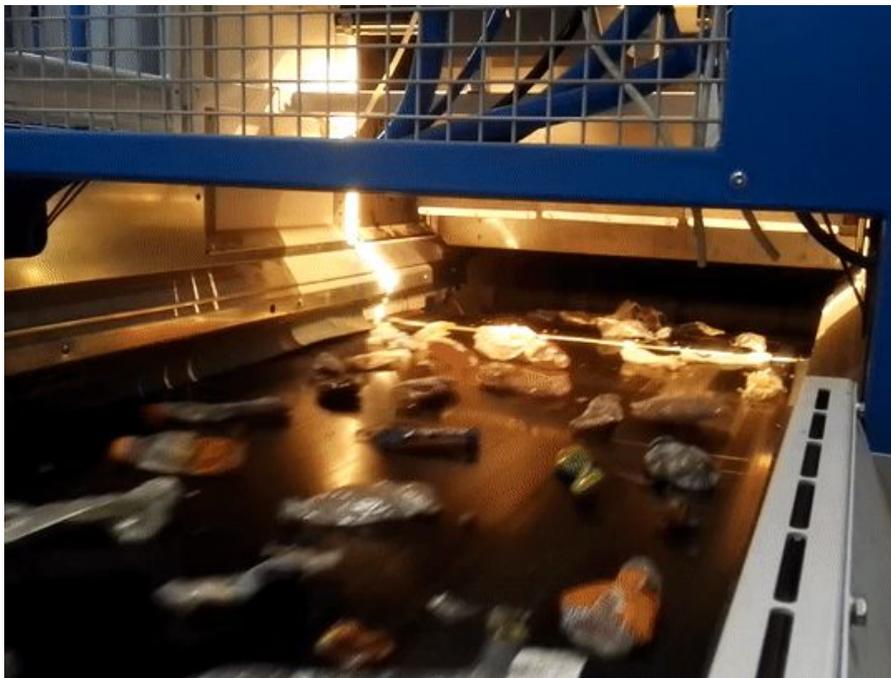
Frameworks & Outils MLOps pour gérer le cycle de vie des projets d'IA

Use Cases & AI in Action

Tri de déchets plastiques automatisé - PELLENC ST



Tri de déchets Inférence de 2ms



Tri de déchets plastiques automatisé - PELLENC ST

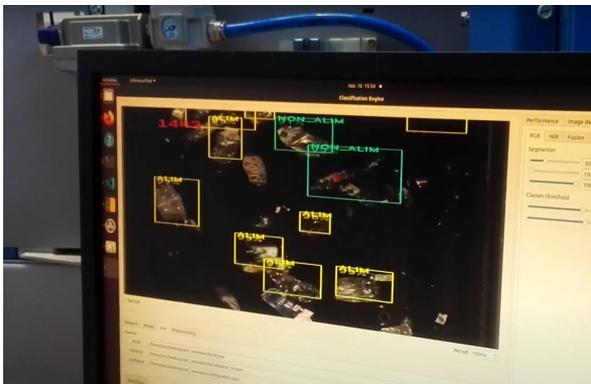


Réalisation

- Technologie de classification de bouteilles PET en utilisant le Deep Learning, adaptée pour fonctionner en temps réel sur des cartes graphiques NVIDIA.
- Collaboration étroite ; de la conception de l'architecture à l'entraînement du modèle, pour mener ce projet de bout en bout.
- Optimisation des réseaux de neurones et l'architecture matérielle pour répondre aux contraintes industrielles de PELLENC ST, en garantissant des temps de traitement inférieurs à 4 ms par image.

Bénéfices

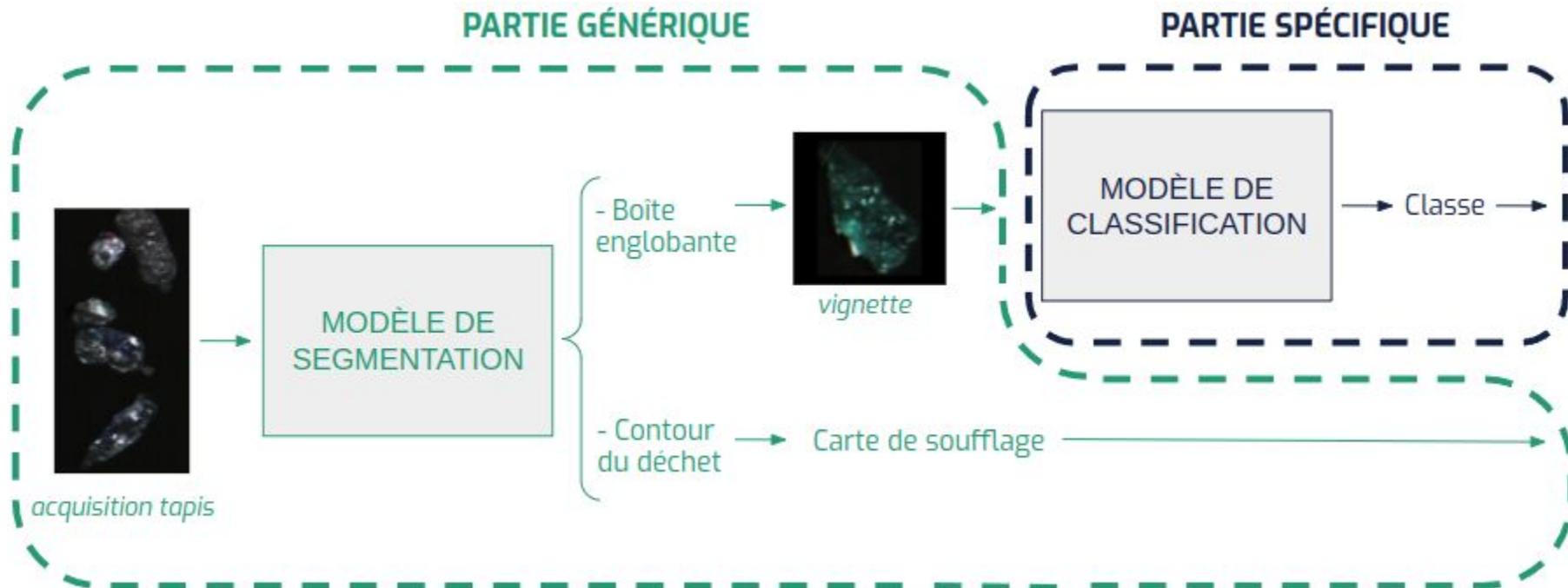
- PELLENC ST dispose désormais d'une technologie de pointe qui permet de classifier en temps réel des bouteilles en plastique PET.
- Capacité de trier les déchets de manière très précise et efficace, ce qui se traduit par une réduction des erreurs de tri, une augmentation de la fiabilité du processus et des gains environnementaux.
- Traitement plus de 1000 bouteilles à la seconde.



Technologies & expertises associées

- Computer vision, deep learning, on premise, C++, GPU

Pipeline & chaîne de traitement multi-modèles



Ce découplage permet d'envisager un modèle de segmentation robuste et générique et des modèles de classification spécifiques entraînés pour une tâche précise.



Détection de défauts sur des Wafers de Silicium



Les apports de Neovision

- Deep Learning, IA Générative & Segmentation Sémantique d'images: solution innovante à l'état de l'art pour répondre aux contraintes de jeux de données limités via la génération de données simulées
- Fine-tuning pour permettre d'annoter les défauts de manière automatique
- Optimisation mémoire pour entraîner une IA sur des images de très grande dimension sans perte de précision.

Les bénéfices pour STMicroelectronics

- Automatisation de la détection de certains défauts critiques
- Amélioration de la qualité des produits de STMicroelectronics
- Augmentation de l'efficacité de la production

Détection de défauts sur des Wafers de Silicium

Réalisation

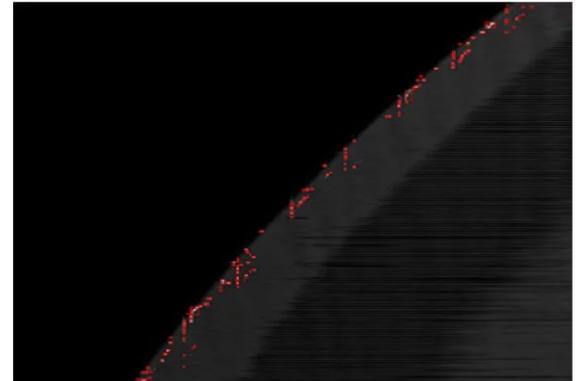
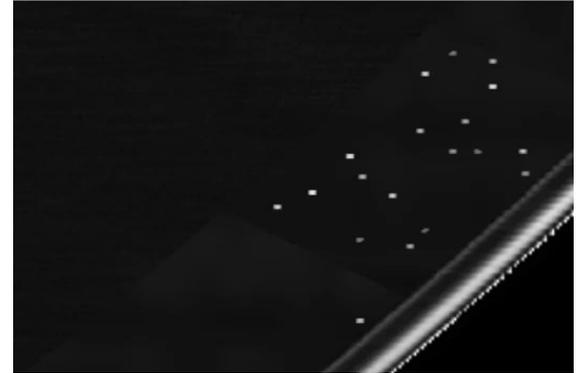
- Utilisation d'un modèle de segmentation sémantique d'images qui a été fine-tuné pour fonctionner sur les défauts de STMicroelectronics et annoter les défauts de manière automatique.
- Mise en place d'une méthode de simulation artificielle de données qui a permis de multiplier par dix la taille de la base d'entraînement.
- Sélection d'une fonction de coût adaptée pour la segmentation et des méthodes d'optimisation stochastique ont été utilisées pour trouver les meilleurs paramètres du modèle.

Bénéfices

- Développement d'une IA de détection de défauts robuste et performante
- Amélioration du savoir-faire de équipes STMicroelectronics et de la performance de son processus de contrôle qualité.
- Automatisation de la détection de certains défauts critiques se traduisant par une amélioration de la qualité des produits de STMicroelectronics et une augmentation de l'efficacité de sa production.

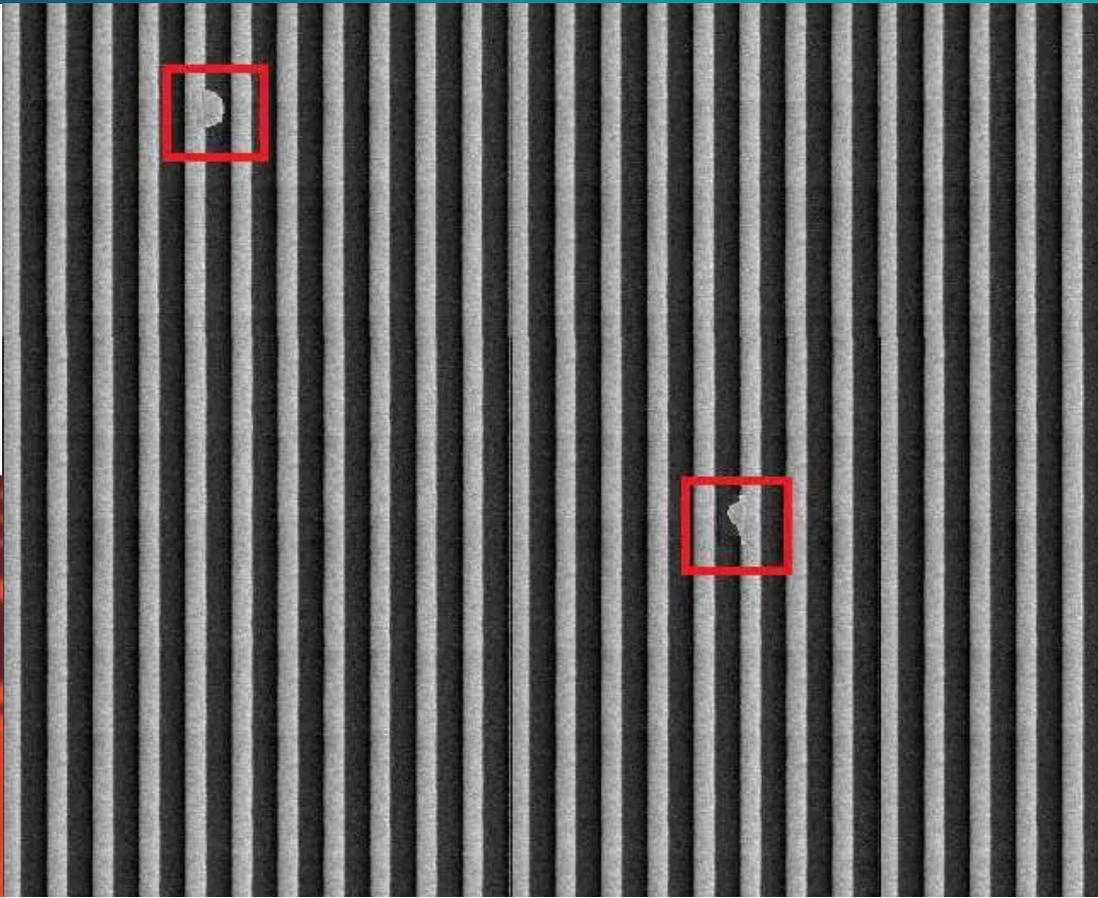
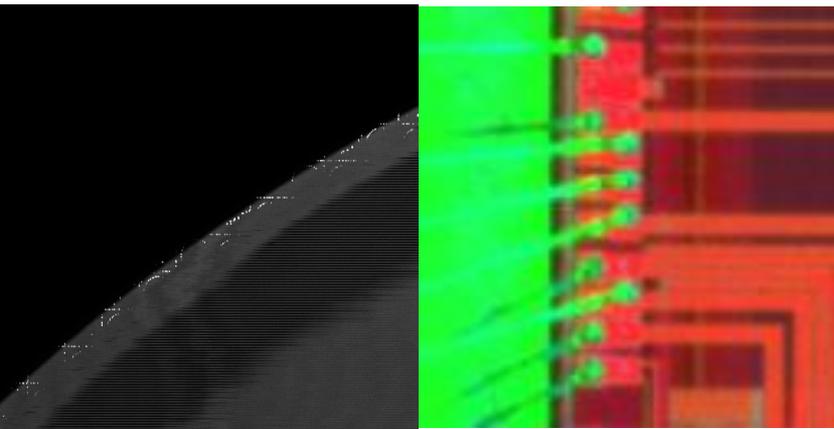
Technologies et expertises associées

- Génération de données, IA Générative, Segmentation d'image, Deep learning,



La défectivité : pourquoi détecter des défauts dans un semi-conducteur ?

- ↘ **Performances** du semi-conducteur
- ↘ **Rendement** de la ligne de production
- ↘ **Répétabilité** du procédé de fabrication
- ↗ Impact sur la **miniaturisation**



Détection des défauts verriers — Tiama

Tiama, un leader mondial des solutions de contrôle qualité, de productivité et de traçabilité pour les industriels du verre.

Jusqu'à récemment la détection des défauts se faisait via des techniques d'analyses d'images classiques. L'IA, le Deep Learning en particulier, ouvre **de nouvelles perspectives** :

- Mieux détecter les défauts,
- Les classifier,
- Offrir une solution plus facile d'utilisation.

Tiama et Neovision ont débuté en 2015 une collaboration de long terme pour progressivement intégrer l'IA dans ses solutions de contrôle du verre.

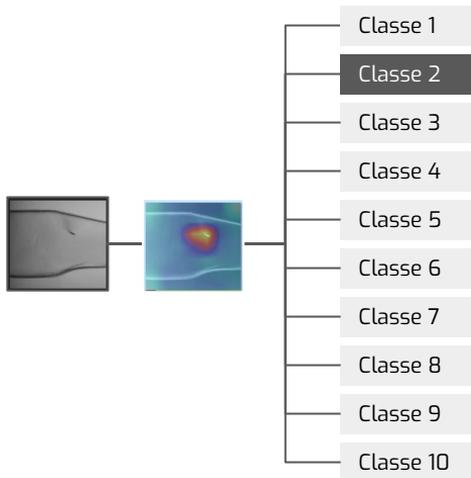


Détection des défauts verriers — Tiamas

Détection



Classification



Les apports de Neovision

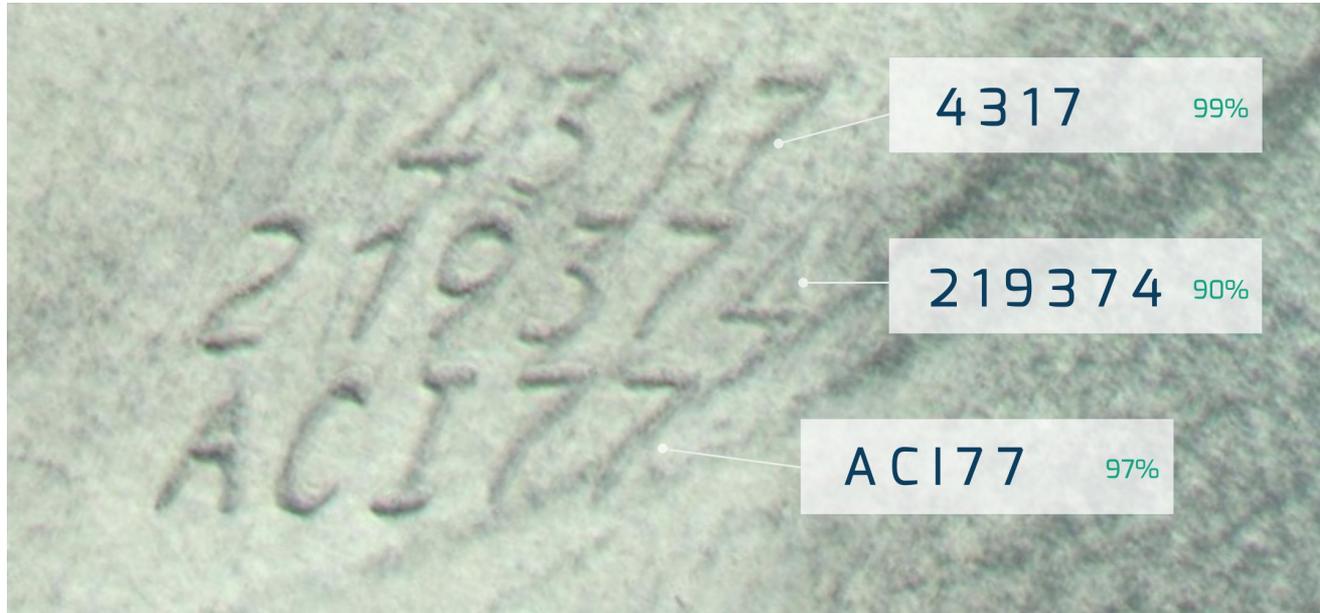
- Du conseil sur l'acquisition et l'annotation des données
- La mise à disposition et la formation des équipes Tiamas à Tadaviz, notre outil d'exploration et d'annotation des jeux d'images
- Une démarche progressive de mise en oeuvre des algorithmes d'IA
- Un transfert de compétences continu aux équipes Tiamas
- Une veille technologique centrée sur les enjeux de Tiamas

Les bénéfices

- Un premier algorithme de classification avec un score F1 moyen $>0,95$ et un délai d'exécution $<50ms$, compatible avec les contraintes de production, intégré à une machine prototype
- Des recommandations sur les besoins d'images supplémentaires et sur les modalités d'annotation
- La montée en compétences des équipes Tiamas

Lecture de caractères gravés dans des peaux— CTC

Une technologie OCR capable de lire les gravures laser présentes sur les peaux de veaux.



OCR sur lecture laser sur du cuir - CTC

Réalisation

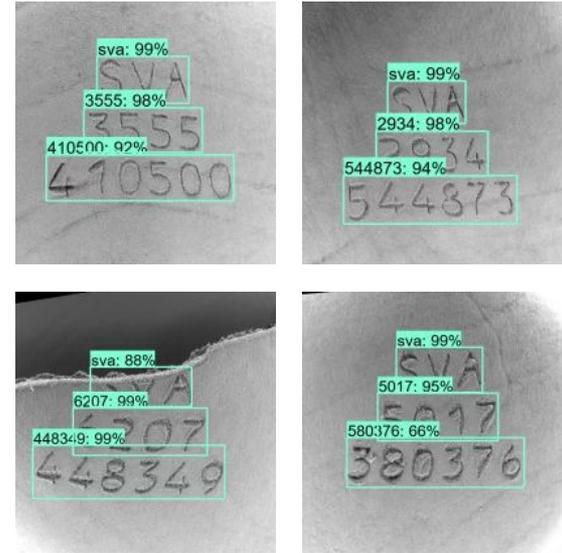
- Accompagnement dans l'acquisition de données exploitables.
- Développement d'une technologie OCR capable de lire automatiquement les gravures laser présentes sur les peaux de veaux, malgré les importantes déformations de caractères.
- Développement d'une solution logicielle complète pour le CTC, mise en production en conditions réelles pour valider ses performances dans plusieurs tanneries.

Bénéfices

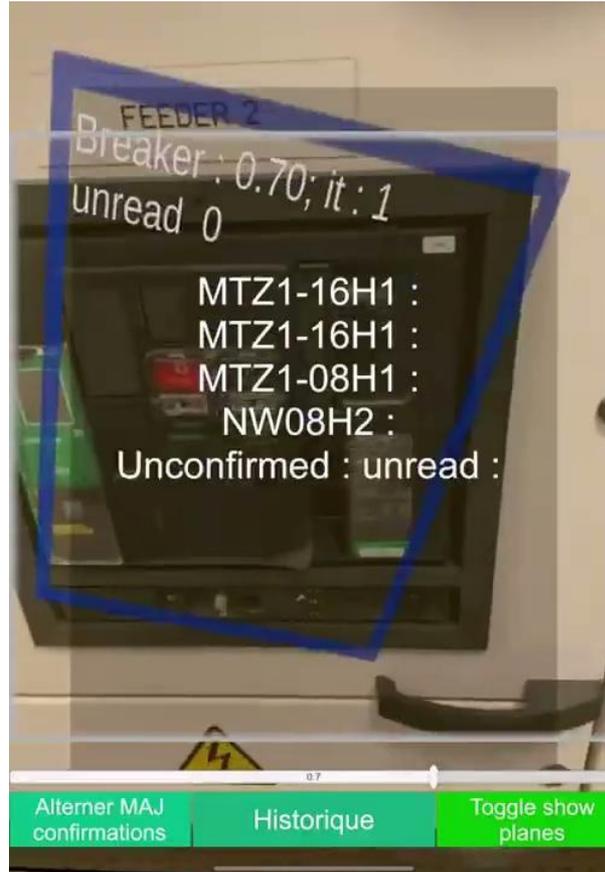
- Automatisation du processus d'identification des peaux de veaux qui accélère le processus de traçabilité.
- Qualité de la gestion des stocks et satisfaction des clients.
- Le CTC envisage de commercialiser cette technologie à l'échelle internationale, ce qui pourrait renforcer sa position de leader et d'expertise dans le secteur, tout en offrant une solution innovante à d'autres acteurs de l'industrie du cuir.

Technologies et expertises associées

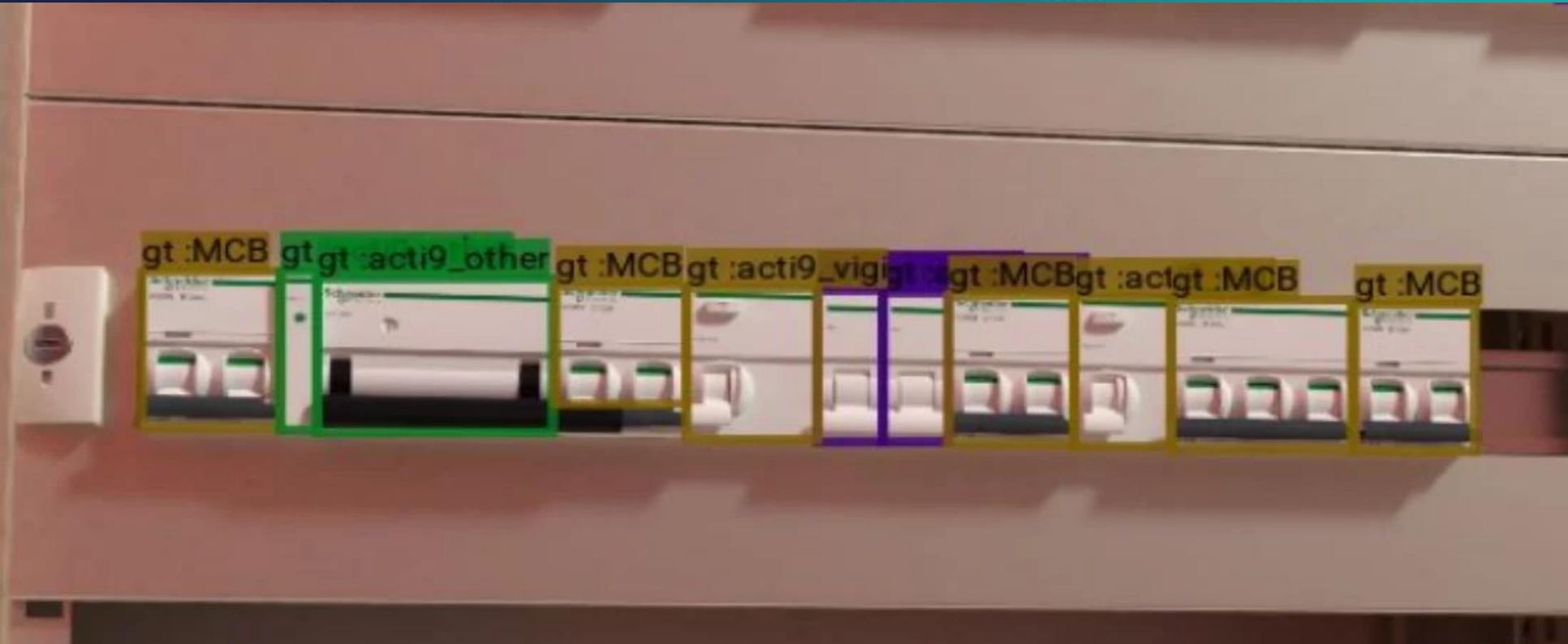
- Computer vision, Deep learning, R&D



IR, OCR, Augmented Reality & Tracking 3D



Détection et reconnaissance de disjoncteurs SCHNEIDER ELECTRIC



Détection et reconnaissance de disjoncteurs SCHNEIDER ELECTRIC



Réalisation

- Application mobile multifonction intégrant des technologies telles que le Deep Learning, la Computer Vision et la Réalité Augmentée et fonctionnant en hors ligne.
- L'application embarque un détecteur d'objets pour localiser les disjoncteurs et une technologie OCR pour lire les références sur les étiquettes des produits.

Bénéfices

- Innovation et expérimentation de l'utilisation de l'IA, de la réalité augmentée et du smartphone pour identifier ses produits.
- L'application permet d'améliorer la gestion des produits en créant des jumeaux numériques, en automatisant la saisie des références et en facilitant la recherche de produits dans le catalogue.
- Les équipes sur le terrain peuvent utiliser l'application sur leur smartphone, sans avoir besoin d'une connexion réseau en temps réel.
- Nombreuses possibilités d'utilisation, notamment la recherche visuelle dans le catalogue produit, une meilleure connaissance du parc de produits installés, la recommandation de produits adaptés.

Technologies et expertises associées

- Deep learning, Computer vision, Edge computing, Smartphone, Réalité augmentée (AR)

3D scan recognition of electrical switchboard



Caractérisation d'activité humaine pour optimiser la sécurité IA multi-modale dans le visible et le non-visible



Sécurité et détection de personnes, véhicules et engins de chantier



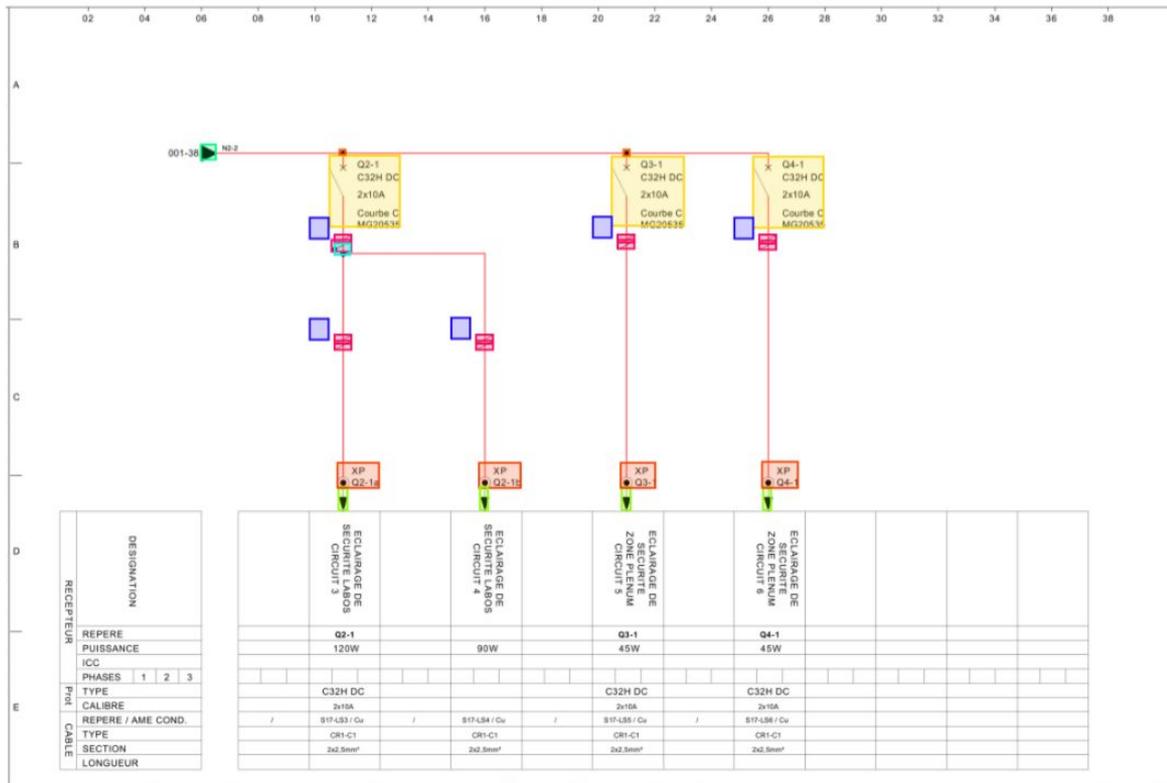
Maintien du niveau de qualité de service dans les aéroports

Comptage de personnes pour l'optimisation du trafic



Optimisation des process qualité & fonctions support

Analyse visuelle de documents techniques



armoire s17_02.png

Name	Count
Bornes	4
CONNEX	2
Disjoncteur	3
Départ	4
Renvoi de folio de type VIENT_DE	1
UNI_1PE	1
UNI_CABLE_I	5
neutre	1
phase	9



Qualité de la supply chain via l'optimisation d'emballage pour les industries pharmaceutiques



Les apports de Neovision

- L'expertise en conception d'algorithmes de deep learning
- Création d'une solution de prédiction de variation de la température intérieure du colis et du nombre d'excursions (lorsque le produit est exposé à une température pouvant dégrader sa qualité)

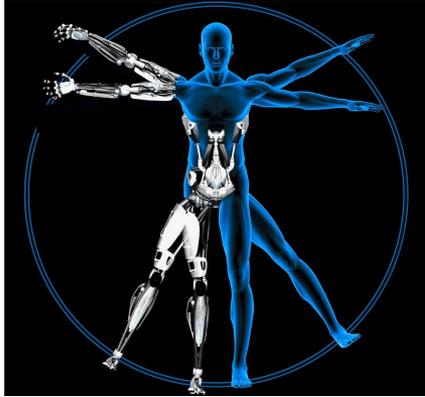
Les bénéfices

- Un choix d'emballage adapté
- Limiter les risques liés à la chaîne du froid
- Optimiser sa chaîne logistique
- Optimiser l'efficacité des produits

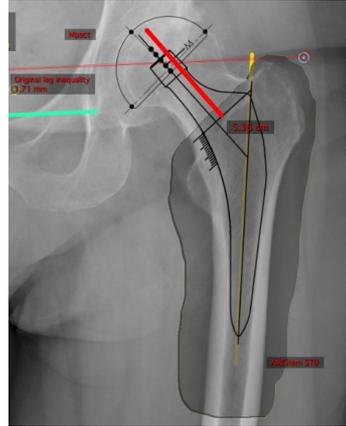


L'IA et la 3D : détection des points d'intérêts

Transferts technologiques issus du médical



Pose de prothèse



Planification chirurgicale

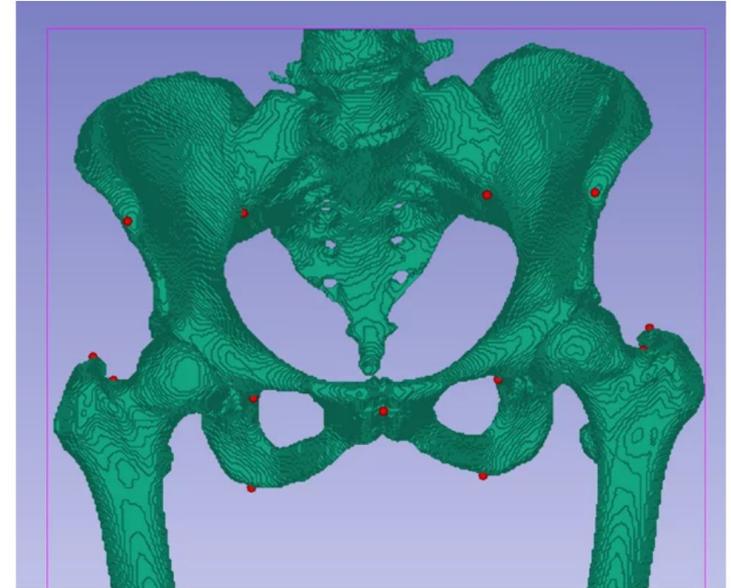


Intelligence Artificielle



L'IA et la 3D : détection des points d'intérêts Transferts technologiques issus du médical

Solution de planification préopératoire qui propose des recommandations pour le type d'intervention, le choix des prothèses et leur implantation.

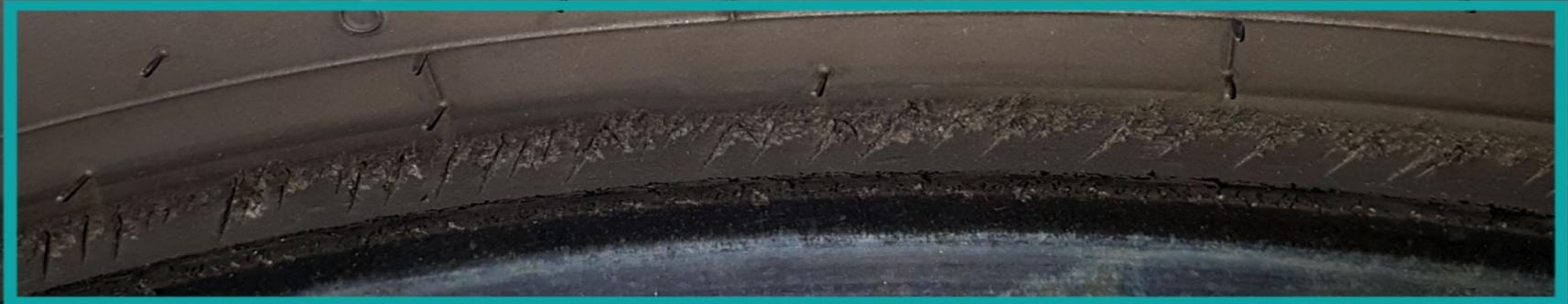


Positionnement des points anatomiques par l'IA pour la pose de prothèse de hanche

L'IA et la 3D : Optimisation du diagnostic de pneus MICHELIN en fin de vie



Reconnaissance de dommages sur des pneus d'avion MICHELIN



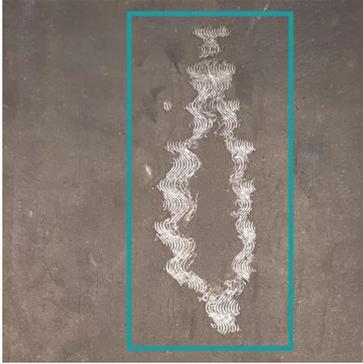
Oblique Bead Cracks

Save

Procedure

Reconnaissance de dommages sur des pneus d'avion

MICHELIN



Réalisation

- Mise en place un protocole d'acquisition d'images représentatives des dommages sur les pneus, puis a annoté ces images pour créer un jeu de données essentiel au développement de l'IA.
- Développement d'une technologie permettant aux techniciens de diagnostiquer le dommage de manière objective via une application smartphone ou tablette le dommage et sa typologie.
- Développement d'une technologie permettant un diagnostic précis et des instructions pour le suivi complet de la durée de vie d'un pneu.

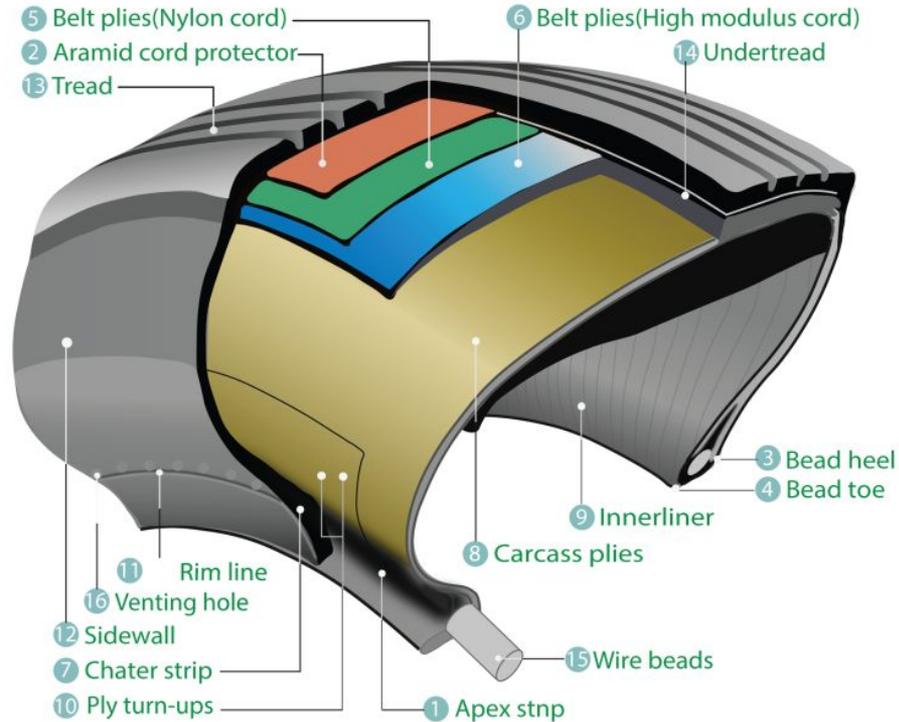
Bénéfices

- Technologie permettant une meilleure efficacité opérationnelle.
- Amélioration de la traçabilité des pneus en enregistrant les dommages
- Procédures précises à suivre, réduisant ainsi les erreurs dans le diagnostic des dommages sur les pneus.

Technologies et expertises associées

- Deep learning, Computer Vision, Classification d'Images, Smartphone

Contrôle qualité industriel



**Comment est réalisé
un projet d'IA ?**

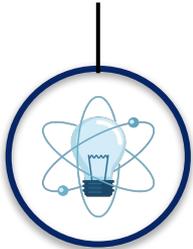
**De l'idéation à
l'Industrialisation**

Méthodologie MLTRL

Un accompagnement global et sur-mesure

Diagnostic / Design :

- Analyse de risque
- Dimensionnement
- Acculturation à l'IA



Processus itératif

POC :
Démonstrateur

Étude :

- Conception
- Développement
- Analyse des performances



Mise en production :

- Intégration Backend
- Portage C++
- Embarqué

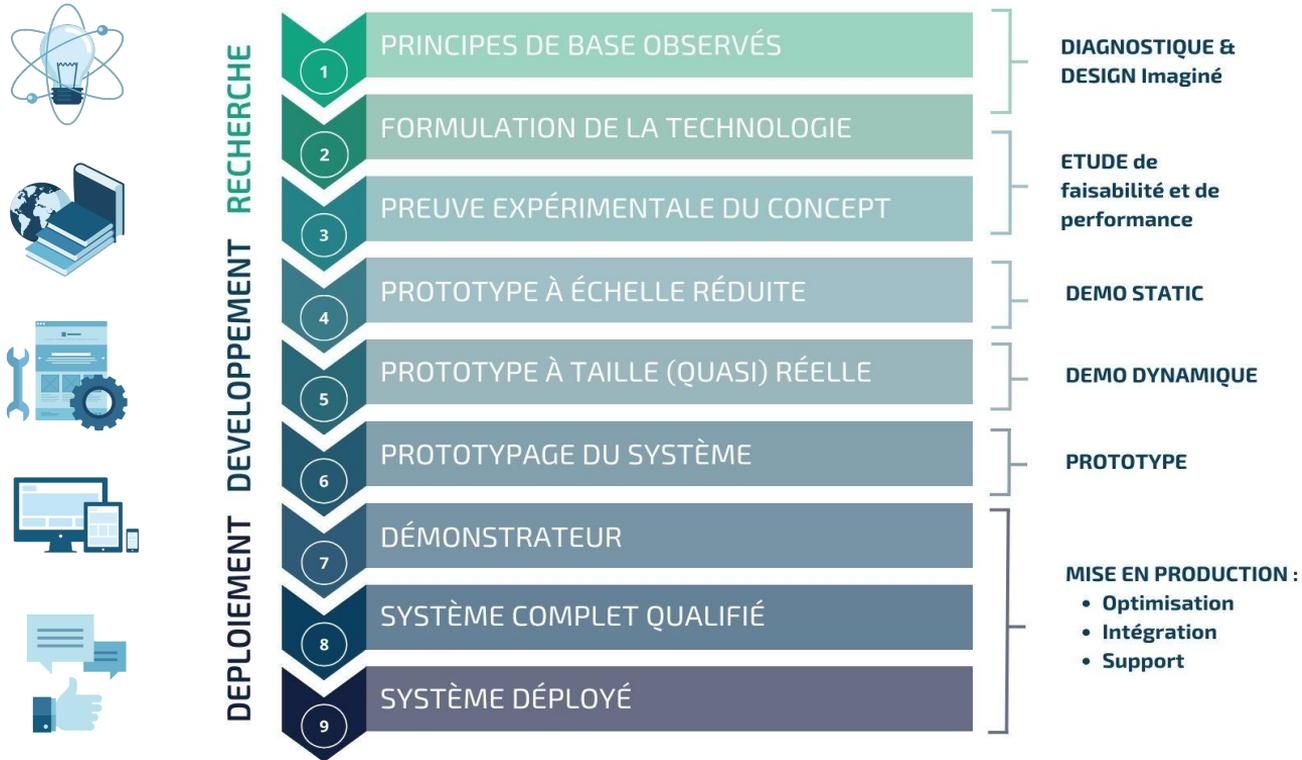


Support :

- Transfert technologique
- Assistance technique



MLTRL - Technology Readiness Levels for Machine Learning systems



**Les Données, la clé
d'une IA efficace**

**& l'Optimisation des
modèles d'IA**

Les données

La clé d'une IA efficace

1

Collecte de données

Pour entraîner des modèles d'IA, il faut des **quantités massives de données** organisées de manière à être comprises par les architectures d'IA.

2

Traitement des données et temps réel

Les données doivent pouvoir circuler rapidement entre la périphérie, les data centers et le cloud pour permettre un traitement en temps réel.

3

Mise à jour continue : DataOps & MLOps

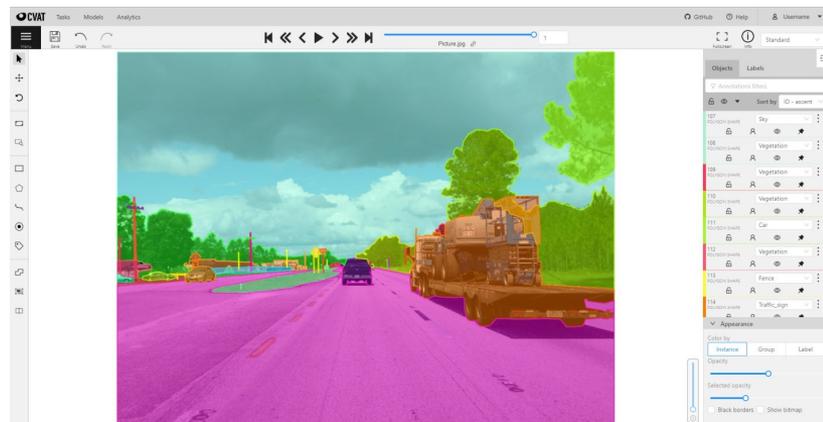
Les modèles ré-entraînés avec de nouvelles données sont renvoyés aux périphériques pour toujours plus de précision.



CVAT - Outil d'annotation de la vérité terrain

CVAT est un outil d'**annotation de données** qui permet :

- L'annotation d'**images** et de **vidéos**,
- Le traitement de tâches de **détection**, de **classification** et de **segmentation** d'images
- L'annotation rapide grâce à une interface **web ergonomique** (interpolation, pré-annotation)
- La **sécurisation** de vos données par [https/login/mdp](https://login/mdp)
- Le travail **collaboratif**

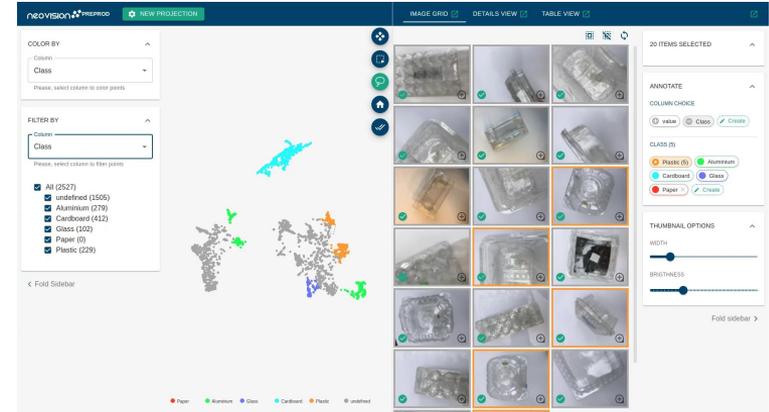


CVAT

Tadaviz - notre outil pour gérer la qualité de vos datasets

Tadaviz est un outil de **data mining** qui permet :

- D'avoir une **vision globale** sur ses données
- La **navigation fluide** dans de grandes quantités de données : zoom, déplacement, sélection au lasso
- La recherche de **données similaires**
- Le **contrôler la qualité** : erreurs et biais
- La modification et création **massive** de métadonnées
- Le **suivi de l'évolution** d'un jeu de données



tadaviz

Data cleaning avec TadaViz

Logo: tadaVIZ

COLOR BY

Column

label

Please, select column to color points

ADD A NEW FILTER

< Fold Sidebar

IMAGE GRID 518 ITEMS SELECTED

DETAILS VIEW

The screenshot displays the TadaViz software interface. On the left, a sidebar contains a 'COLOR BY' section with a dropdown menu set to 'label' and a note 'Please, select column to color points'. Below this is an 'ADD A NEW FILTER' button and a 'Fold Sidebar' option. The main area on the left shows a scatter plot of data points, with a central cluster of points colored in various colors (green, blue, red, purple) and a dashed white circle highlighting a specific region. On the right, an 'IMAGE GRID' shows 518 items, each represented by a small image of a plastic bottle or container. Each item in the grid has a green checkmark and a plus sign in the bottom right corner, indicating it is selected. The grid is organized into rows and columns. At the top right, there are navigation icons for home, search, and settings, along with a 'DETAILS VIEW' button. The 'NEOVISION' logo is visible in the bottom right corner.



Optimiser la collecte et les pipelines de données pour l'IA

1

Collecte de données

Collecter et organiser les données nécessaires aux modèles d'IA.

2

Traitement des données

Traiter et préparer les données pour les rendre exploitables par les algorithmes d'IA.

3

Déploiement en production

Déployer les solutions d'IA dans un environnement hybride et multicloud.

TRL 7 à TRL 9 - Pourquoi optimiser les modèles ?

- Besoin en réactivité
 - Latence et débit
 - Passage à l'échelle
- Ressources limitées
 - Mémoire, capacités de calcul
 - Batterie
- Réduction des coûts de service
 - Consommation d'énergie
 - Quantité de matériel d'inférence



Matériel spécialisé avec carte de calcul IA NVIDIA Jetson

De l'entraînement au déploiement



Entraînement



Optimisation



Déploiement



Maintenance

Optimisation
des **métriques**

Export optimisé
pour l'**inférence**

Requêtage sur
un hardware
optimisé

Optimisation pour l'inférence



Objectifs :

- Réduire la **taille** du réseau (stockage)
- Réduire la consommation de **mémoire**
- Améliorer la **latence** et le **débit** d'inférence

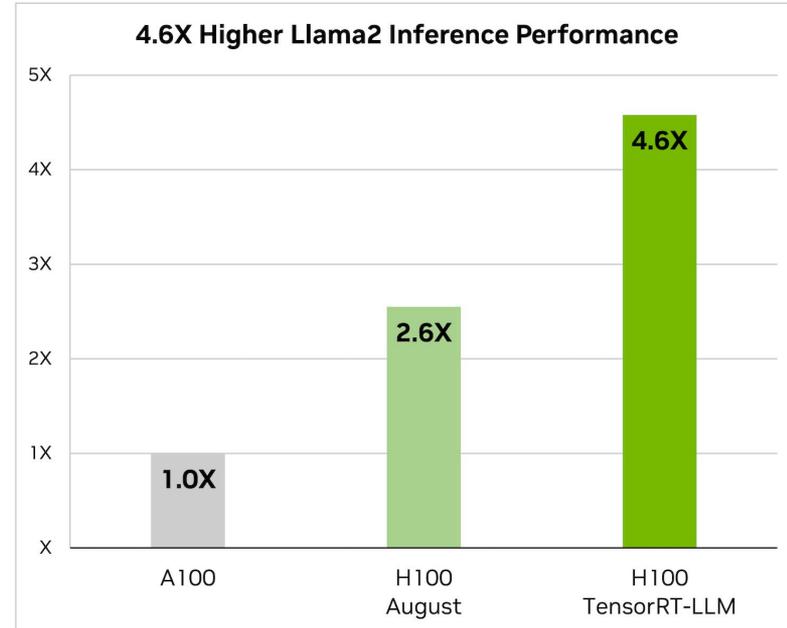
➤ Avec une perte minimale de la précision !

Frameworks & Outils MLOps pour gérer le cycle de vie des projets d'IA

Frameworks d'optimisation

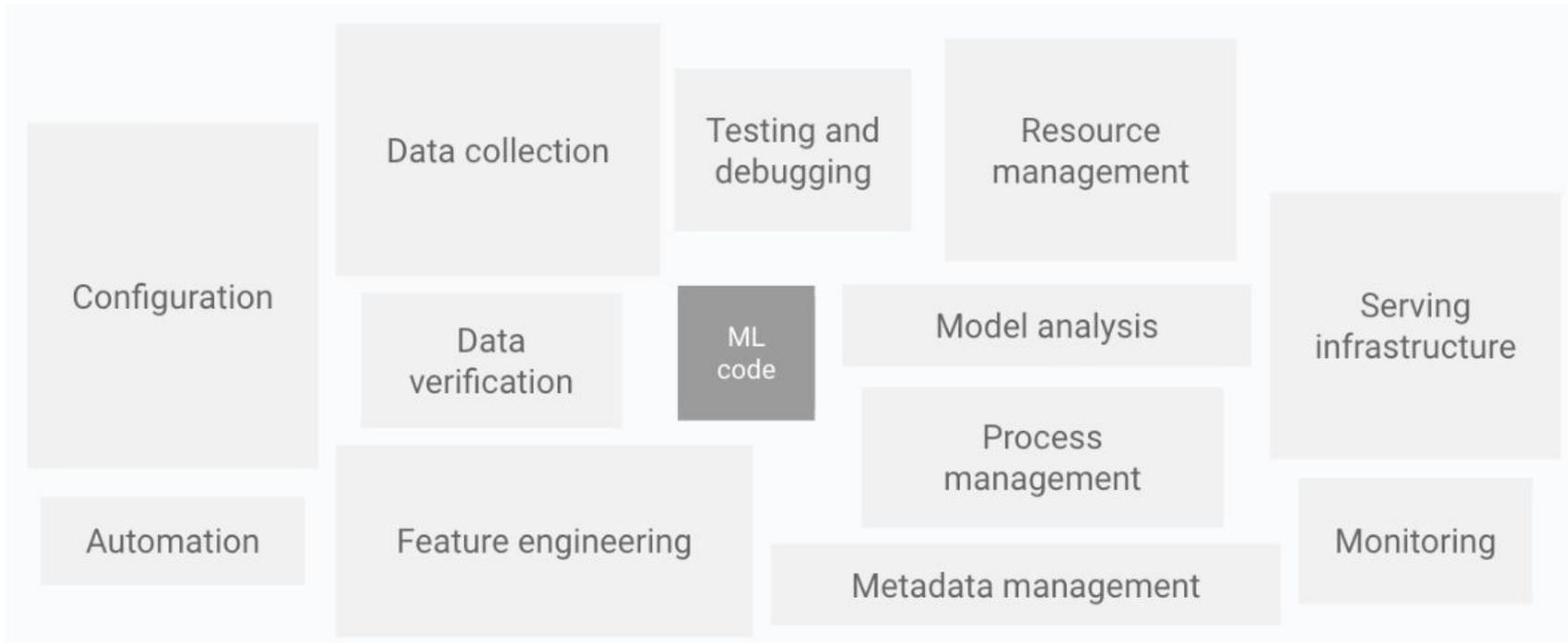
- TensorRT et TensorRT-LLM
 - Serveurs et cartes de calculs NVIDIA
- ONNX
 - Multi-plateformes
- CoreML et TFLite
 - Appareils iOS/macOS et Android
- OpenVINO
 - Matériel Intel
- Etc.

➤ **Les gains peuvent être importants :
x2, x10, x100 selon le modèle**

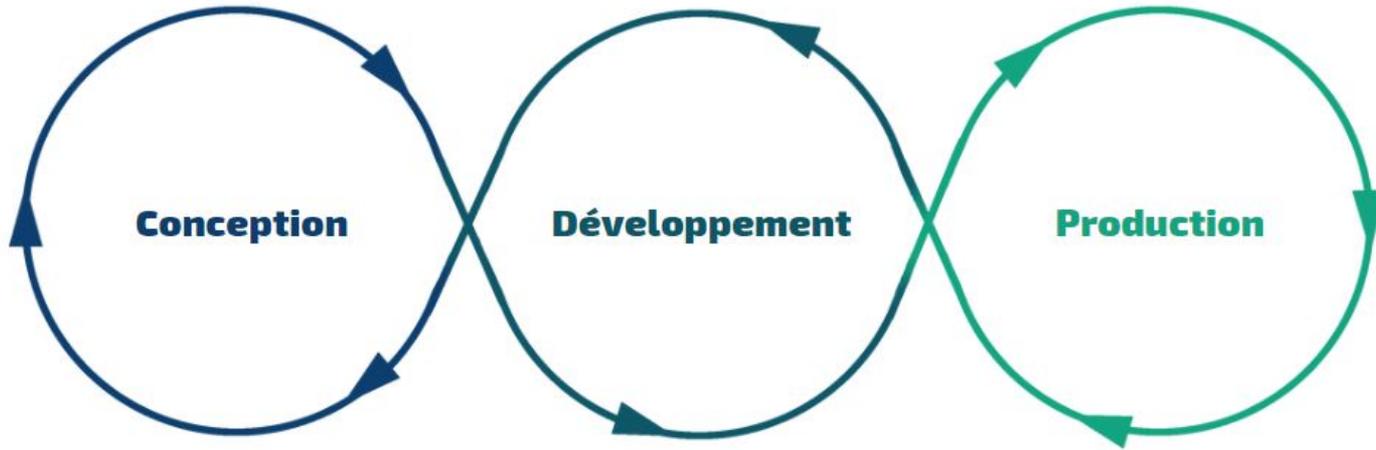


Exemple de gains de performance apportés par l'optimisation d'inférence TensorRT-LLM

DevOps & MLOps



MLOps

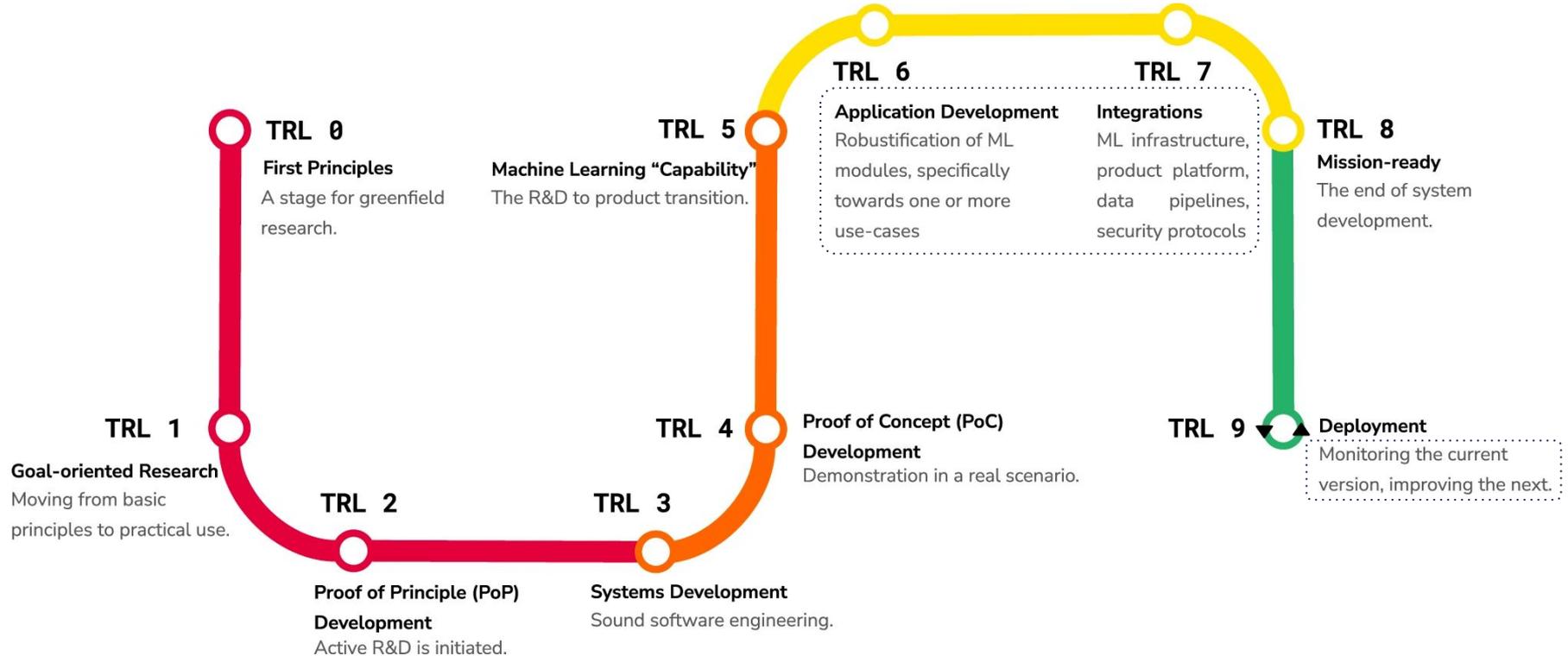


- Priorisation des cas d'usages
- Ingénierie des exigences
- Validation des données

- Ingénierie des données
- Ingénierie des modèles
- Test et validation des modèles

- Déploiement des modèles
- Intégration & Déploiement continue
- Suivis et déclenchement

DevOps, MLOps & MLTRL



Différents niveaux de MLOps



Niveau 1

- Obtenir des **résultats comparables** et **reproductibles**
- **Gérer les versions** des modèles, des datasets et des résultats
- Un dataset de référence pour **benchmarker** les modèles



Niveau 2

- Lancer automatiquement un **entraînement**
- **Déployer** automatiquement un modèle
- **Détecter des erreurs** dans le dataset, le pipeline ou le déploiement



Niveau 3

- **Monitorer** les performances d'un modèle
- Détecter **data drift**, concept drift et anomalies
- Automatiser les **tests**

[Technology readiness levels for machine learning systems, Nature Communications](#), Oct. 2022

Rôle de Neovision dans l'écosystème

L'IA Accessible à Tous

Accompagnement End-to-End du Design Thinking à l'Industrialisation

2014
Depuis

20+
Expert.es

150+ Clients
400+ Missions

60+
Use cases
en Production

Equipes composées d'**Ingénieurs logiciels**,
d'Ingénieurs IA et de **Docteurs** en IA

Expertise en **algorithmie** et en **Data Science**

Des **services IA** pour créer des IA
sur-mesure

Des **outils data** pour créer des IA
performantes

Veille constante des dernières avancées
Développe des technologies à **l'état de l'art**



IA Embarquée



Deep Learning



Data Prep



Multimodal

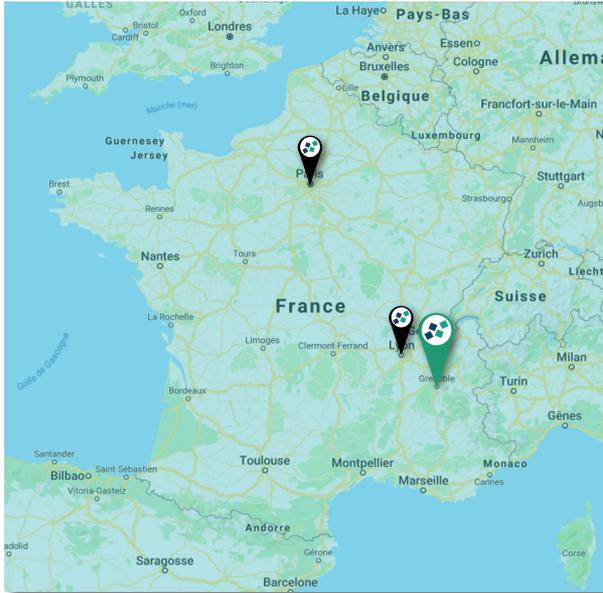


Computer Vision



LLMs

Notre implantation, nos marchés & nos clients



Numérique



Environnement



Santé



Industrie



Energie



Smart City

23%



Start-up

47%



PME & ETI

30%



Grand groupe

Un accompagnement multi-sectoriel



Notre positionnement charnière au sein de l'écosystème



IA Booster France 2030

Un dispositif en 4 phases pour aider les PME et ETI à valoriser leurs données et à intégrer des solutions d'intelligence artificielle

A travers le déploiement de dispositifs de sensibilisation à l'IA (phase 1) et de modules de conseil individuel (phases 2 à 4), IA Booster France 2030 vise à accompagner les PME et ETI françaises dans leurs processus de transformation numérique, d'enrichissement de leur offre par de nouveaux services, d'amélioration de leur compétitivité et de modernisation de leur appareil de production grâce à l'intégration de solutions d'intelligence artificielle.

	1 SENSIBILISATION ACCULTURATION	2 DIAGNOSTIC DATA IA	3 CHOIX DE L'APPROCHE IA	4 EXPÉRIMENTATION DE LA SOLUTION IA
Objectifs	<ul style="list-style-type: none"> Accélérer la prise de conscience des dirigeants d'entreprise de l'intérêt de l'adoption de l'IA Favoriser la montée en compétence des dirigeants sur le sujet 	<ul style="list-style-type: none"> Evaluer les pistes de développement à partir de l'exploitation des données Cadrer les conditions de mise en œuvre d'un projet stratégique de rupture : développement d'un nouveau business model, de nouveaux produits/services Identifier un ou plusieurs cas d'usage et définir un plan d'action 	<ul style="list-style-type: none"> Réaliser l'étude de faisabilité des solutions IA pertinentes au regard des besoins identifiés Fournir des éléments d'aide à la décision entre développement interne et recours à une solution disponible sur le marché Établir le plan de mise en œuvre et évaluer risques et impacts 	<ul style="list-style-type: none"> Être accompagné dans le déploiement opérationnel de la solution IA identifiée
Modalités	<ul style="list-style-type: none"> Parcours de e-learning Autodiagnostic IA Webinaires Évènements / communication 	<ul style="list-style-type: none"> Diagnostic Data Intelligence Artificielle (10 J/H) 	<ul style="list-style-type: none"> Mission de 10 J/H 	<ul style="list-style-type: none"> Financement de prestations intellectuelles uniquement Modalités de déploiement à définir
Cibles	<ul style="list-style-type: none"> Toutes entreprises, avec une priorité aux PME et ETI 	<ul style="list-style-type: none"> PME et ETI françaises de tout secteur En priorité, effectifs compris entre 10 et 2000 et CA annuel supérieur à 1 000 000 € Ayant réalisé les phases précédentes ou non 		<ul style="list-style-type: none"> PME & ETI en sortie de phase 3
Coût et taux de subvention	<ul style="list-style-type: none"> Gratuit 	<ul style="list-style-type: none"> 13 000€ HT subventionné à 50% 	<ul style="list-style-type: none"> 13 000€ HT, subventionnés à 50 % 	<ul style="list-style-type: none"> 60 000 € HT maximum (sur devis), subventionnés à 50%
Prix entreprise	<ul style="list-style-type: none"> Gratuit 	<ul style="list-style-type: none"> 6 500€ HT maximum 	<ul style="list-style-type: none"> 6 500€ HT 	<ul style="list-style-type: none"> 50% du montant

Livres Blancs

NetApp

neovision

LIVRE BLANC

Deep Learning et
Computer Vision,
un duo technologique
stratégique pour votre
organisation



**HUB
FRANCE
IA**

**CHOISIR UN MODELE
D'IA GENERATIVE
POUR SON ORGANISATION**

Jun 2024

LLMs, RAG, modèles génériques et modèles spécifiques

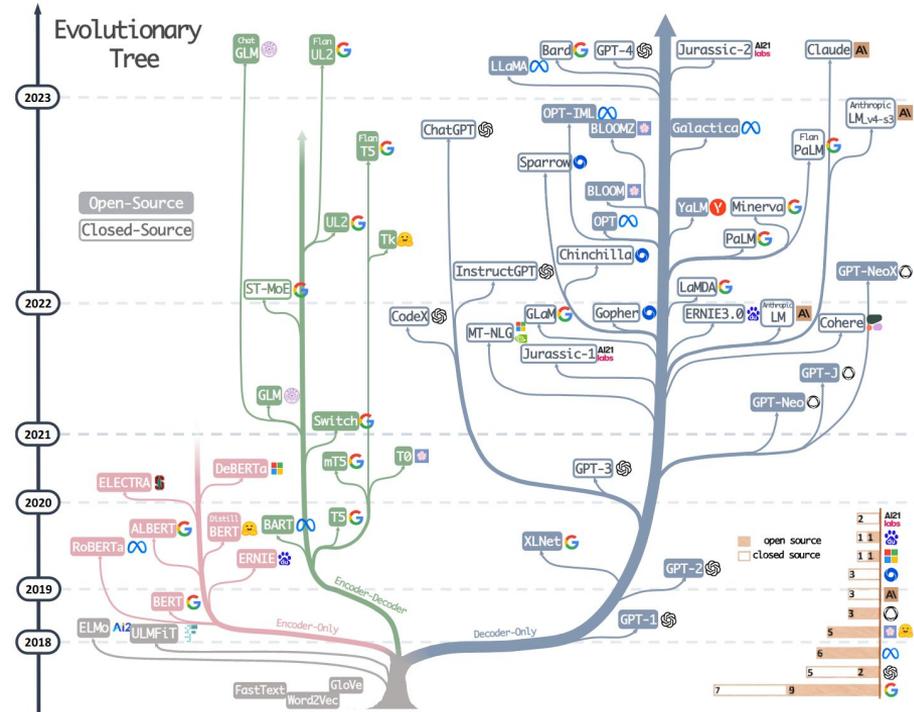
Comment s'y retrouver dans la myriade de modèles et de versions disponibles sur le marché ?

Expertise IA & compréhension profonde des Architectures

Sélection, évaluation, **benchmarks** rigoureux et interprétation des résultats des métriques de **performance** de chaque modèles

Optimisation et **fine-tuning** des modèles pré-entraînés sur des corpus spécifiques, ajustement des hyperparamètres

Ethiques, régulation & **AI Act** : confidentialité des données, biais et transparence



LLMs vs SLMs

Modèles génériques & spécifiques

	ChatGPT 3.5	ChatGPT 4 Turbo	Gemini 1.5 Advanced	Mistral Le chat	Llama 3	Claude 3	Microsoft Copilot
 AppyLearn Appylearny.fr v. 3 - Yvan DEMUMIEUX	 OpenAI ChatGPT gpt-3.5-turbo	 OpenAI ChatGPT 4.0	 Gemini	 MISTRAL AI	 Meta Llama 3	 Claude 3	 COPILOT
Coût	Gratuit	Plus : 20\$/mois Team : 25\$/p/mois	19,99 \$ / mois (Studio gratuit dispo le 02/05/24)	Gratuit (bêta) Payant (entrepri.) Mistral (France)	Gratuit (pas d'accès grand public) Meta	Gratuit (Sonnet) Pro 20\$/m (Opus)	Gratuit Pro : 22 € / mois MS 365 : 30 \$ /m Microsoft (basé sur GPT)
Compagnie	Open AI	Open AI	Google	Mistral (France)	Meta	Anthropic	Microsoft (basé sur GPT)
Date de lancement	30 novembre 2022	6 novembre 2023	15 février 2024	26 février 2024	18 avril 2024	03 mars 2024	7 février 2023
Recense des données	02/2022	12/2023	11/2023 En ligne	01/2021	Inconnu	08/2023	04/2023 En ligne
Max Tokens (contexte en entrée)	4 096 (3000 mots)	8 000 / 16 000 (6000 / 12000)	De 128 000 à 1 000 000 (700 000 mots)	32 000 (20 000 mots)	8 000 (6000 mots)	200 000 (150 000 mots)	4000 (caractères)
API	Turbo : 0.5 / 1.5	Turbo : 10 / 30 4 : 30 / 60 32K : 60 / 120	1.0 Pro : 0.5 / 1.5 1.5 : 7 / 21	8x7B : 0.7 / 0.7 Med. : 2.7 / 8. Large : 8 / 24	8Bx70B Model (Open Source)	Haiku : 0.3 / 1.25 Sonnet : 3 / 15 Opus : 15 / 75	
Cout\$/ million Tokens In/Out							
Disponibilité	164 pays	164 pays	180 pays	Inconnue	12 pays non FR	EU, USA...non FR	Inconnue
Accessibilité	Webapp	Webapp + iOS + Android	Webapp + iOS + Android	Webapp	Appli Metas ou multILLM (Poe...)	Webap	Webapp + iOS + Android
Langue de l'interface	FR	FR	FR	FR	EN	EN	FR
Rédaction FR (classement)	5 ^e	 1 ^e	 2 ^e	4 ^e	?	3 ^e	?
Vitesse / Perf (classements)	3 ^e	4 ^e	 1 ^e	3 ^e	?	1 ^{er} ?	?
Test de QJ*	6 ^e (64)	2 ^e (85)	5 ^e (77.5)	Non testé	Non testé	 1 ^{er} (101)	4 ^e (79)

Les modèles génériques (**LLMs** >10B) sont capables de résoudre des tâches générales et dotés d'une forte capacité de raisonnement

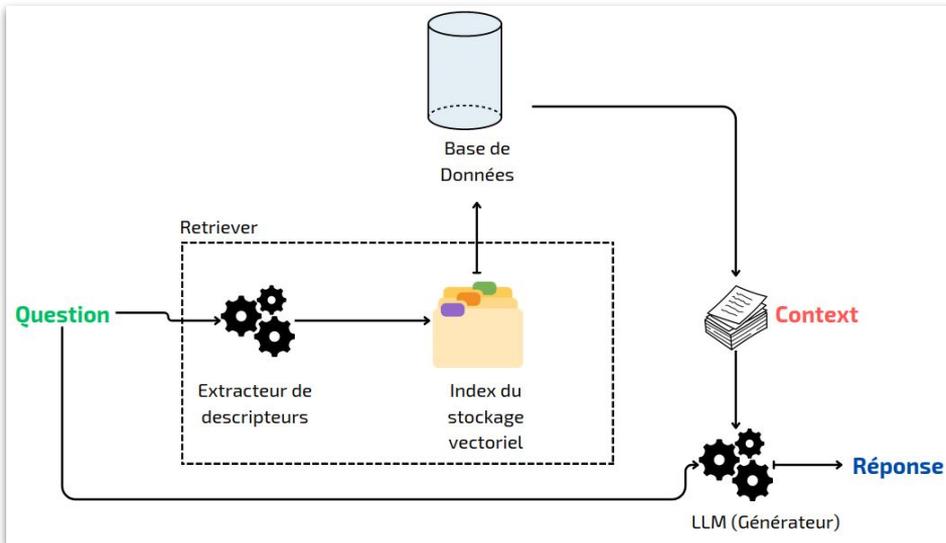


- limités par les données de pré-entraînement
- ne connaissent pas les événements récents

Fine-tuning de modèles spécifiques (**SLMs** <10B), petits LLMs, sur des datasets et corpus documentaires spécifiques afin d'optimiser leurs performances et leur coût

Les RAG

Reinforcement Learning with Human Feedback



Accès à des informations à jour et spécifiques

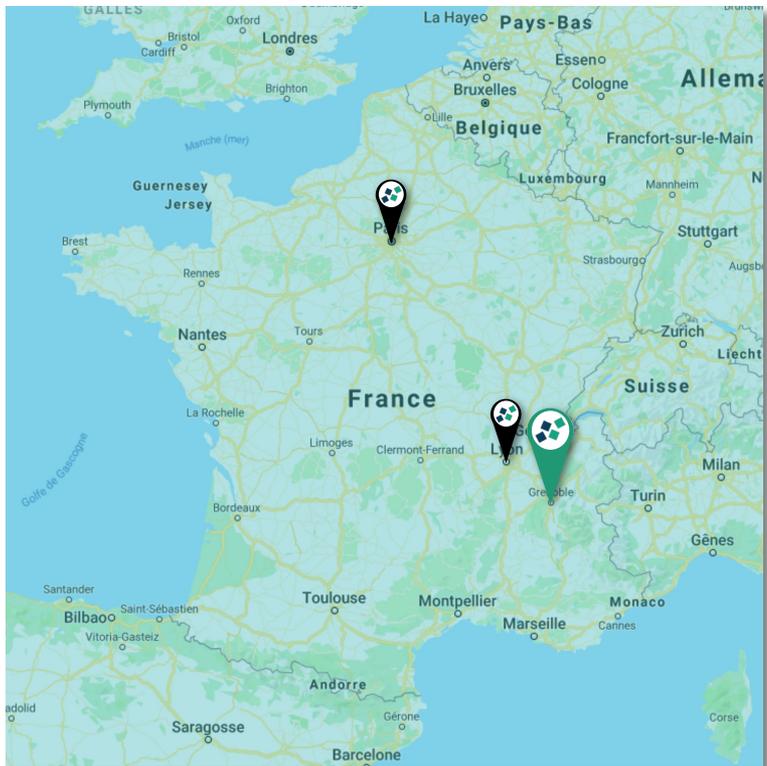
Les LLMs sont formés sur un large corpus de données préexistantes mais peuvent ne pas contenir des informations spécifiques ou à jour. En intégrant un composant de récupération, le modèle peut accéder à des informations externes en temps réel.

Contextualisation et personnalisation

Les **RAG** (Retrieval-Augmented Generation) sont des modèles de NLP et permettent de :

- tirer le meilleur parti des connaissances préexistantes
- améliorer la performance en intégrant de la **contextualisation**, de la **personnalisation**, de la **veille** et des feedbacks et bases de données métier

Merci pour votre attention



Gaël Ricard

Partners & Business Development

+33 7.83.02.38.77

gael.ricard@neovision.fr

www.neovision.fr

Pour plus d'informations sur l'IA, suivez-nous !



@Neovision SAS



@neovisionSAS



@neovision