

Traçabilité et Identification des IAs et des contenus générés par IAs

Mihai Mitrea



Quiz : cerner le vraie du faux !



Problème : cerner le vraie du faux ?



GÉNÉRÉES PAR GPT4.0

Nécessité de
réguler l'IA

AI ACT



Traçabilité
(recital 27)

Registration
(recital 53)

Suivi
(article 12)

(1) *Entrée en vigueur progressive dès 2024*
(2) *Confusion totale au niveau de solutions techniques*

- Comment les utilisateurs et les distributeurs d'IA peuvent se conformer à l'AI ACT ?



Contexte

- Les technologies d'IA générative se démocratisent :



- Leur évolution est rapide vers des contenus de plus en plus en plus réalistes :



mars 2023

avril 2025

Défis majeurs : désinformation, confiance de l'audience/l'utilisateur, plagiat, droits d'auteurs
Objectifs : transparence, intégrité et IPR

Contexte

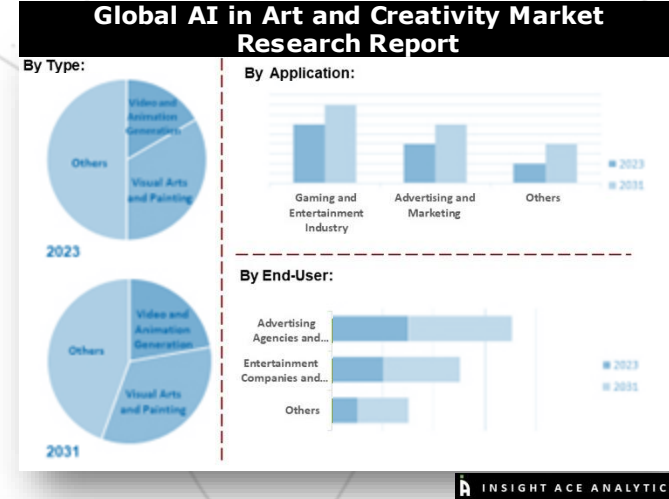
Generative AI in Creative Industries
Market to Reach \$21.6 Billion, by
2032 at 29.6% CAGR: Allied Market
Research



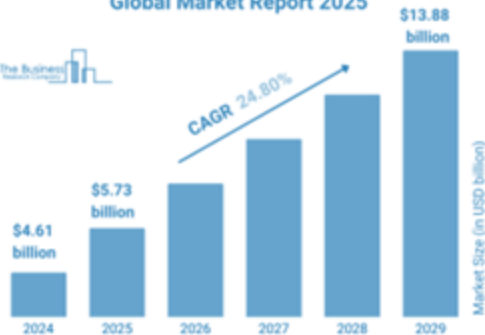
AI-generated art sales increased by 600% from 2020 to 2023



45% of music producers have used AI to compose or enhance tracks



Artificial Intelligence (AI) In Art and Creativity Global Market Report 2025



Gitnux

- 78% of users believe AI tools improve their creative productivity
- AI-powered generative design reduced project timelines by an average of 40% in architecture firms
- 56% of graphic designers reported increased efficiency after adopting AI tools
- 72% of fashion brands have experimented with AI for trend forecasting

Contexte

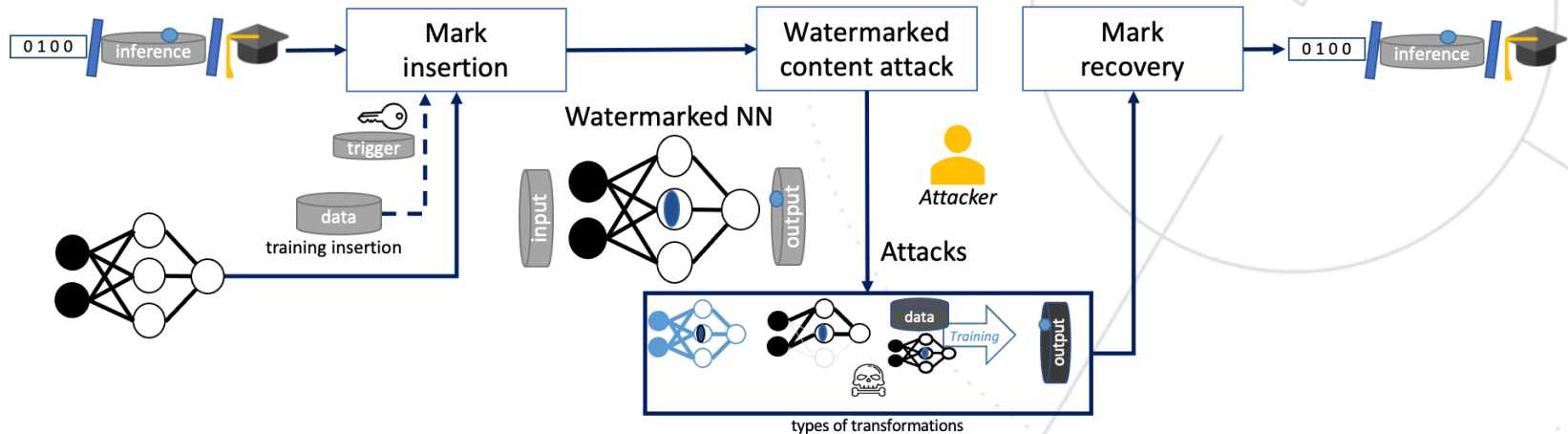
- **Tracer et identifier à la fois les IAs en elles-mêmes et les contenus générés par les IAs**
- **L'intelligence artificielle est à la fois un outil de production et le résultat d'un processus de production**
- **L'intelligence artificielle devient transverse par rapport à la verticale métier:**
 - industries créatives et culturelles, médias, industrie 4.0, informatique, télécommunications,...
- **L'intelligence artificielle comporte une duplicité de principe:**
 - un logiciel qui produit une sortie pour une entrée
 - un ensemble de représentations binaires qui accomplissent une même tâche



Source: ChatGPT

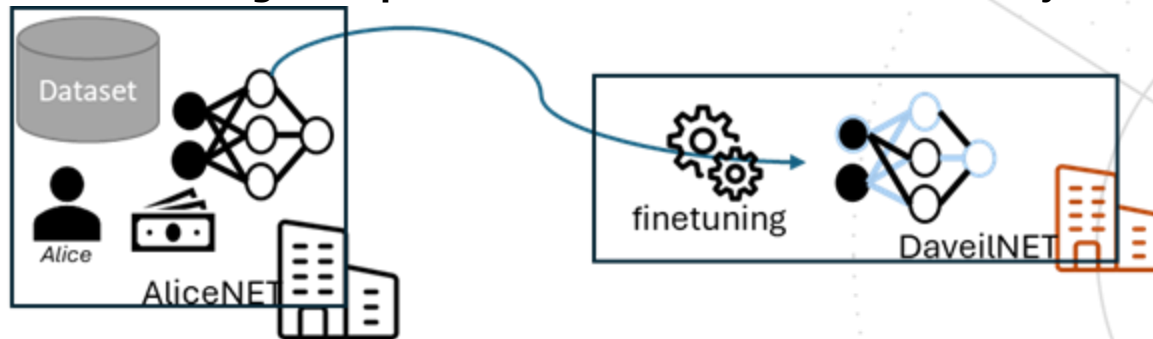
Tatoauge pour les IAs

- Une famille des outils méthodologiques et applicatives pour insérer des données supplémentaires (des marques) d'une façon imperceptible et persistante, selon une clef secrète



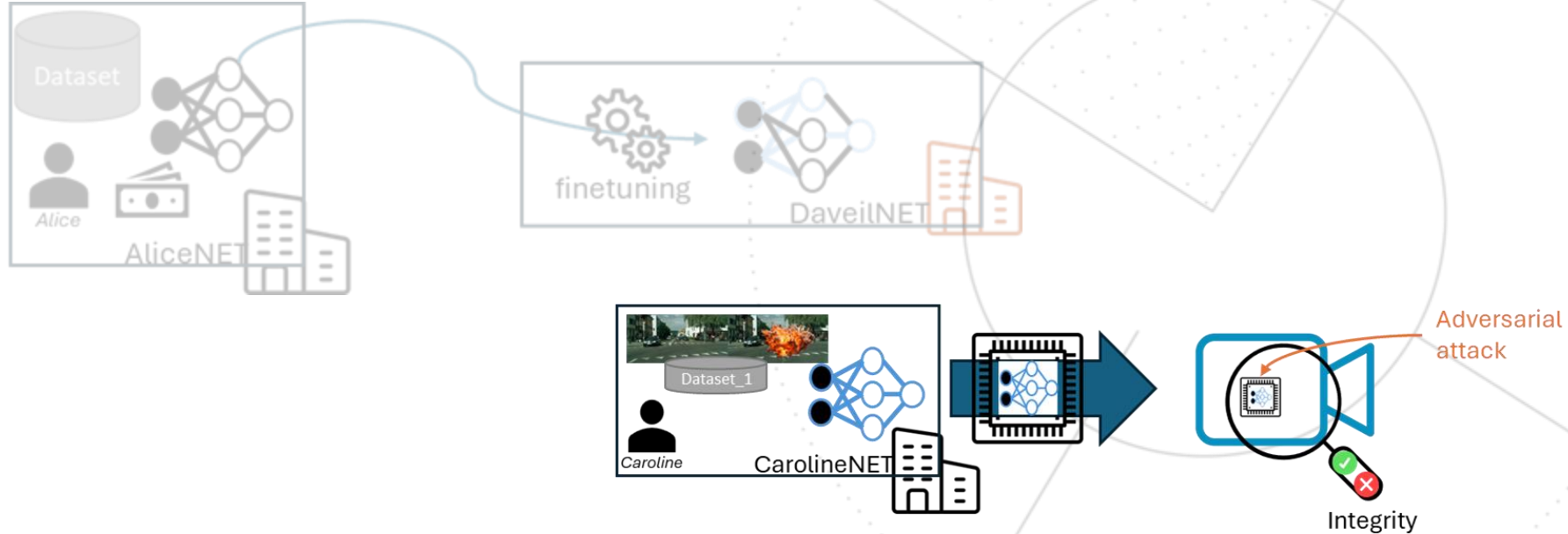
Bénéfices applicatives ciblées

- Trois cas d'usage complémentaires : détection de contrefaçon de l'IA



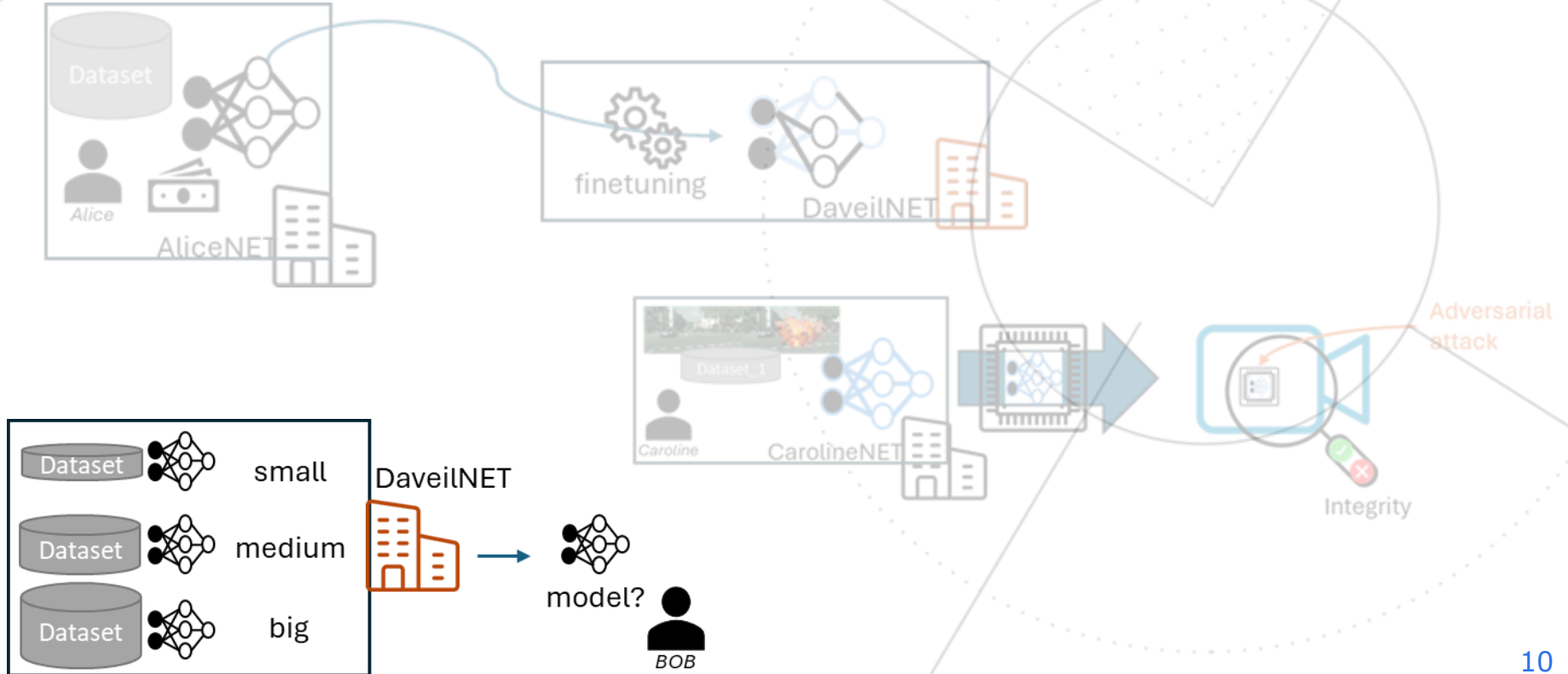
Bénéfices applicatives ciblées

- Trois cas d'usage complémentaires : contrôle de l'intégrité de l'IA



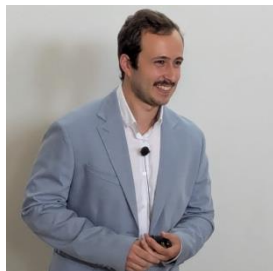
Bénéfices applicatives ciblées

- Trois cas d'usage complémentaires : détection de QoS de l'IA



Notre solution: le tatouage

- L'équipe



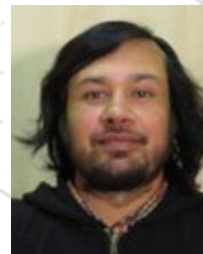
Carl De Sousa Trias



Mihai Mitrea



Enzo Tartaglione
Expert en IA



Sumanta Chaudhuri
Expert en IA embarquée



Marco Cagnazzo
Expert en compression

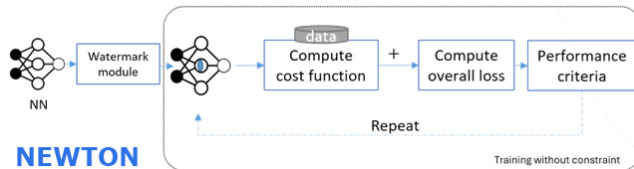
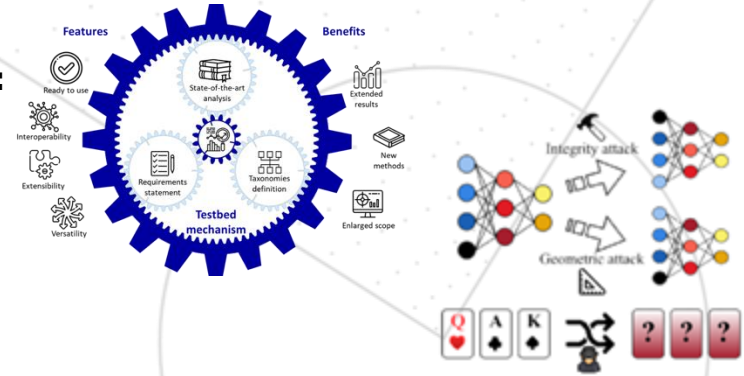


Attilio Fiandrotti
Expert en compression d'IA

Socle méthodologique et applicatif NEWTON (1/2)

• Etat d'avancement & résultats obtenus (2021-2025):

- Etude de l'état de l'art
 - Création d'un *testbed* open-source de référence
 - Création de taxonomies complètes pour le tatouage
- Contributions théoriques, méthodologiques et techniques
 - Etude des attaques malveillantes de types géométriques
 - Etudes des méthodes de tatouage RNA substitutive
 - Spécification de 2 normes internationale IEEE-3304 et IEEE-3309
- Invention d'une nouvelle méthode de tatouage
 - Spécification algorithmique finalisée
 - Déclaration INPI FR2505074 mai 2024
 - Implémentation logicielle finalisée

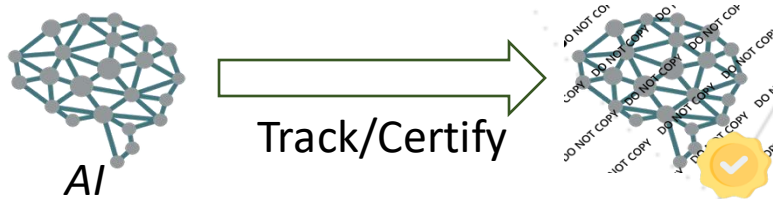


NEWTON

	Methodological properties				Working properties			
	multi-bit payload	Imperceptibility	Robustness	Computational complexity	Flexibility	Scalability tinyML	Versatility	Genericity
[Uchida17]	✗	✓	✓	?	?	?	✗	✓
[Rouhani19]	✗	✓	✓	?	✗	?	✗	✗
[Li21]	✓	✓	✓	?	?	?	✗	✓
[Yao24]	✗	✓	✓	?	✓	?	✗	✓
[Tartaglione21]	?	✓	✓	✗	?	✗	✗	✓
[Lv23]	✗	✓	✓	✗	✗	✗	✗	✓
[DeSousa24-2]	?	✓	✓	✗	?	✗	✗	✓
[Tond24]	✓	✓	✓	?	?	?	✗	✓
[Botta21]	✗	✓	✗	✓	✗	?	✗	✓
[Kakikura22]	✗	✓	✓	✓	✗	?	✗	✓
[Bazyad24]	✗	✓	✓	✓	✗	?	✗	✓
[Zhang18]	✗	✓	✓	?	✗	?	✗	✗
[Adil18]	✗	✓	✓	?	✗	✓	✗	✗
[Wen23]	✗	✓	✓	?	✗	?	✗	✗
[Luo25]	✗	✓	✓	✗	✗	?	✗	✗
NEWTON	✓	✓	✓	✓	✓	✓	✓	✓
NEWTON-UP	✓	✓	✓	✓	✓	✓	✓	✓

Technologie NEWTON (2/2)

- NEWTON permet de tatouer les IA et leurs inférences:
 - **Versatilité:** Le moment de l'insertion (avant/pendant/après) de l'IA
 - **Imperceptibilité:** Aucun impact sur les performances du modèle
 - **Scalabilité:** Un faible coût d'exécution (insertion et detection)



Le désordre informationnel

- **Projet DemoCIS : Démocratie & Désordre informationnel a pour objectifs :**
 - Repenser la citoyenneté et la démocratie face à la fragmentation des espaces publics
 - Créer une Convention citoyenne et un Baromètre de la démocratie
 - Développer des outils et formations pour chercheurs, citoyens et institutions
- 4 défis majeurs :
 - Innovation démocratique et renouvellement des décisions publiques
 - Repenser les formes d'expression citoyenne
 - **Agir contre la polarisation et les désordres informationnels**
 - Renforcer les démocraties menacées
- Porté par:

Christine Balagué

Professeure qualifiée des Universités en Sciences de Gestion Institut Mines-Télécom (IMT-BS)
Membre du comité d'éthique de la Défense et de l'advisory board de l'Accelerator de Princeton University sur l'environnement de l'information



Le désordre informationnel

• Recherche:

- Production de connaissances sur les désordres informationnels et leur écosystème
- Effet de la régulation (DSA)
- Focus sur l'efficacité des contre mesures

• Technologies innovantes:

- Algorithmes de fact checking
- Technologies de watermark
- Large Language Models
- ...

• Action contre la désinformation:

- Fresque de la désinformation
- Formations certifiantes
- Actions étudiants sur les campus et dans les territoires



Réseau Good in tech

Le désordre informationnel

Projet ClimateSense

- Problématique :
 - La **désinformation climatique** (intentionnelle ou non) affaiblit la confiance dans la science, réduit le soutien public à l'action climatique et freine la coopération internationale
- Enjeux majeurs :
 - Complexité scientifique du climat → propice à la manipulation
 - Intérêts politiques et économiques puissants
 - Diffusion massive sur les réseaux sociaux (21% en 2023 selon statista)
- Porté par :



Raphael Troncy

Associate Professor in Data Science at Eurecom (IMT), Sophia Antipolis

Expertise en Knowledge Graphs, Ontologies, NLP, Recommender Systems, Multimedia Semantics



Kaunas University of Technologies (KTU),
Lithuania



Prague University of Economics and
Business (VSE), Czech Republic

Rejoignez-nous !

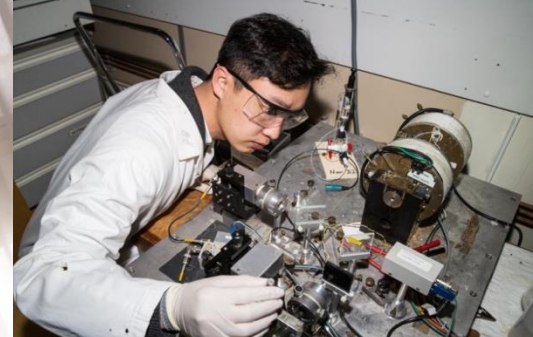
16 mai 2025 au 16 juin 2027

Appel à projets : « Transition numérique de la Culture et appropriation de l'intelligence artificielle »

L'appel à projets « Transition numérique de la Culture et appropriation de l'intelligence artificielle » vise à accélérer la transition numérique des secteurs culturels et créatifs, à travers le développement, l'appropriation et le déploiement de technologies innovantes telle que l'intelligence artificielle, afin de renforcer la compétitivité et la résilience de la filière des industries culturelles et créatives françaises.

- Le **développement de nouveaux outils** contribuant à la production de nouveaux contenus culturels ou aux nouvelles expériences culturelles et artistiques, fondés notamment sur l'IA ;
- Le **développement de solutions de partage et de valorisation de données** ;
- Le **développement de solutions permettant de sécuriser l'exploitation des contenus** et la juste rémunération des ayants droit.





Merci de votre attention

mihai.mitrea@telecom-sudparis.eu

carl.de-sousa-trias@telecom-sudparis.eu

