جامعة ابن طفيل
ⵜⴰⵙⴷⴰⵡⵉⵜ ⵙⵉ ⵎⵓⵃⴰⵎⵎⴷ
Ibn Tofaïl University

Faculté des Sciences

---

**Université Ibn Tofail**

**Faculté des Sciences, Kénitra**

**Mémoire de Projet de Fin d'Etudes**

**Master Intelligence Artificielle et Réalité Virtuelle**

---

# ANIMO : Génération d'Images Artistiques Conditionnée par les Émotions pour une Expérience Thérapeutique Personnalisée

---


**Établissement d'accueil : Le laboratoire CIAD**

Elaboré par : Mlle. NADIA EL MOUDEN

Encadré par : Mr. TARIK BOUJIHA (ENSA KENITRA UIT)
Mr. MOHAMED KAS (CIAD-Lab)


***Soutenu le 26 septembre 2024, devant le jury composé de :***

- Mme. Raja Touahni FS KENITRA (UIT)
- Mr. Anass Nouri FS KENITRA (UIT)
- Mr. Tarik Boujiha ENSA KENITRA (UIT)
- Mr. Idriss Moumen FS KENITRA (UIT)

---

Année universitaire 2023/2024


CIAD
Connaissance et Intelligence Artificielle Distribuées

# Dédicace

**À mes parents,**

Il est difficile de trouver les mots pour exprimer toute la gratitude que je ressens envers vous. Depuis mon enfance, vous m'avez soutenue avec un amour inconditionnel, guidée avec sagesse et encouragée à poursuivre mes rêves sans relâche. Vos sacrifices, votre patience, et votre foi en moi m'ont permis de devenir la personne que je suis aujourd'hui. Merci pour chaque sourire, chaque mot réconfortant, et chaque geste d'amour. Vous êtes et resterez toujours ma plus grande source de force.

**À mes grands frères,**

Vous avez été et continuez toujours d'être mes modèles et mes piliers. Eh bien, chaque fois que je me remets en question, vous arrivez toujours à dénicher les mots appropriés pour me solidifier et encourager à avancer. Votre support constant, des conseils avisés et votre digne présence m'ont toujours donné cet élan de surmonter les victoires de l'existence pour viser toujours plus haut et chercher l'excellence. Merci de m'avoir cru en moi et vous m'avoir inspiré à donner en retour le meilleur de moi-même dans tout ce que je fais, personnellement et professionnellement.

**À ma grande famille,**

Vous avez tous, à votre manière, contribué à mon parcours. Votre amour, vos encouragements, et votre soutien ont été des moteurs puissants qui m'ont permis de franchir chaque étape de ce chemin. Merci pour votre présence et pour les valeurs que vous m'avez transmises, qui m'ont guidée tout au long de cette aventure.

**À mes professeurs et encadrants,**

Je voudrais prendre ce moment pour exprimer ma gratitude à vous tous pour votre dévouement, votre expertise et votre soutien inestimable. Vous m'avez non seulement donné de votre temps pour partager vos connaissances, mais vous avez également éveillé en moi une passion pour l'apprentissage et l'innovation. Que vous soyez mes professeurs d'université ou mes encadrants, que ce soit dans mon établissement d'attache ou au laboratoire, vous avez été des mentors exceptionnels. Votre guidance, vos conseils avisés et les outils fournis m'ont conduite à ma prospérité dans ce projet. Merci d'avoir joué un rôle majeur dans ma formation.

**À mes amies,**

Votre amitié a été une source de joie et de réconfort tout au long de cette aventure. Merci pour les moments partagés, pour vos encouragements constants, et pour votre soutien inébranlable. Vous avez enrichi cette expérience de manière inestimable, et je suis chanceuse de vous avoir à mes côtés.

À tous ceux qui m'ont soutenue, merci du fond du cœur.


**Au nom d'Allah le miséricordieux,**

Avant tout, je tiens à exprimer ma profonde gratitude à Allah, le Tout-Puissant, pour m'avoir accordé la force, la patience et la persévérance nécessaires à la réalisation de ce projet final.

Je tiens également à remercier chaleureusement tous ceux qui ont contribué, directement ou indirectement, à la réussite de ce projet. Je suis particulièrement reconnaissante **à M. Yassine Ruichek** et **M. Mohamed Kas**, mes superviseurs au Laboratoire CIAD (Connaissance et Intelligence Artificielle Distribuées). Leur confiance, leur expérience et leur esprit d'équipe ont créé un environnement propice non seulement à l'apprentissage mais aussi à l'épanouissement personnel et professionnel.

Ma sincère gratitude va également à **M. Tarik Boujiha**, mon conseiller académique, dont les conseils inestimables et les commentaires constructifs ont grandement enrichi ce projet. Ce fut un honneur de travailler sous le mentorat d'un professeur aussi dévoué et engagé.

Je tiens à remercier l'ensemble du corps enseignant du programme de maîtrise en intelligence artificielle et réalité virtuelle, en particulier notre coordinatrice, **Mme Raja Touahni**, ainsi que **Mme Khaoula Boukir**, **M. Anass Nouri** et **M. Messoussi Rochdi**, pour leur soutien indéfectible et leur dévouement à notre éducation.

Je remercie également les membres du jury d'avoir accepté d'évaluer mon travail, ainsi que l'ensemble du corps professoral du Master de l'Université Ibn Tofail de Kénitra, pour la qualité de l'enseignement que j'ai reçu. Enfin, je tiens à exprimer ma gratitude à tous ceux qui m'ont soutenu tout au long de ces cinq années d'études et qui, d'une manière ou d'une autre, ont contribué à la réussite de ce projet.

# Résumé

Ce rapport constitue une synthèse des travaux effectués dans le cadre de projet de fin d'études réalisé au sein du laboratoire de Connaissance et Intelligence Artificielle Distribuée (CIAD), en vue de l'obtention du diplôme de Master en Intelligence Artificielle et Réalité Virtuelle. Le projet, propose une approche novatrice pour l'art-thérapie en utilisant l'intelligence artificielle générative pour transcender les limites des méthodes traditionnelles, souvent jugées insuffisantes pour retranscrire avec précision les émotions humaines. Réalisé en collaboration avec l'artiste française Lina KHEI, ce projet repose sur des modèles spécifiques adaptés à chaque émotion — joie, tristesse, peur, neutralité, colère, surprise, et dégoût — pour générer des images artistiques en fonction de l'état émotionnel détecté chez l'utilisateur.

La problématique abordée dans ce projet est de comprendre comment l'intelligence artificielle peut être utilisée pour créer des œuvres artistiques en fonction des émotions humaines, et d'explorer les impacts potentiels de cette interaction sur le bien-être des utilisateurs. L'objectif principal est de développer un système capable de détecter en temps réel les émotions à partir des expressions faciales et de générer des images qui reflètent ces émotions, offrant ainsi une forme d'expression visuelle personnalisée. Les questions de recherche explorent comment les modèles de diffusion stable (Stable Diffusion) et les modèles LoRA peuvent être appliqués pour adapter les images générées aux émotions détectées ?

Pour atteindre ces objectifs, une approche basée sur la détection des émotions à travers un détecteur de points de repère faciaux a été mise en place. Les émotions détectées sont ensuite utilisées pour sélectionner le modèle LoRA approprié, lequel est chargé pour ajuster la génération d'image. Les techniques de génération d'images utilisent la technologie de diffusion stable, combinée à des modèles LoRA pour affiner les résultats en fonction des émotions détectées. Le projet a également mis en place des mécanismes de gestion de la concurrence pour assurer une détection et une génération d'images fluides, en évitant les conflits dans le processus.

Les principales découvertes du projet ANIMO montrent que l'utilisation de l'intelligence artificielle dans la création artistique émotionnelle est non seulement faisable, mais aussi potentiellement bénéfique pour les utilisateurs en termes de bien-être psychologique. Le système développé a démontré sa capacité à générer des images qui sont perçues comme étant en adéquation avec les émotions détectées.

En conclusion, les résultats obtenus montrent que l'application ANIMO peut offrir une nouvelle forme d'interaction entre l'art et l'utilisateur, où les émotions jouent un rôle central dans la création artistique. Ces découvertes suggèrent des applications potentielles dans le domaine de la thérapie par l'art et d'autres contextes où l'expression émotionnelle est cruciale. Les recommandations incluent la poursuite des recherches sur l'intégration de modèles plus complexes pour gérer des émotions multiples et l'exploration de nouvelles techniques d'amélioration de la fluidité et de la précision de la détection d'émotions.

**Mots-clés :** Intelligence Artificielle, Stable Diffusion, LoRa, Détection des Émotions, Art Thérapeutique, Application Web, Génération d'Images.

# Abstract

This report provides a summary of the work conducted as part of my final project within the Distributed Knowledge and Artificial Intelligence Laboratory, towards obtaining a master's degree in Artificial Intelligence and Virtual Reality. The project introduces an innovative approach to art therapy by leveraging generative artificial intelligence to surpass the limitations of traditional methods, which often fail to accurately capture human emotions. Developed in collaboration with French artist Lina KHEI, the project utilizes specific models for each emotion—joy, sadness, fear, neutrality, anger, surprise, and disgust to generate artistic images based on the user's detected emotional state.

The core challenge addressed in this project is to explore how artificial intelligence can be harnessed to create artistic works that reflect human emotions and to investigate the potential impact of this interaction on users' well-being. The primary objective is to develop a system capable of real-time emotion detection from facial expressions and generate images that mirror these emotions, offering a personalized visual expression. The research questions explore how Stable Diffusion models and LoRa models can be applied to tailor the generated images to the detected emotions.

To achieve these goals, an approach was implemented that detects emotions through a facial landmark detector. The detected emotions are then used to select the appropriate LoRa model, which is loaded to adjust image generation. Image generation techniques utilize Stable Diffusion technology combined with LoRa models to refine the results based on the detected emotions. The project also incorporated concurrency management mechanisms to ensure smooth emotion detection and image generation, avoiding conflicts during the process.

The key findings of the ANIMO project demonstrate that using artificial intelligence in emotional art creation is not only feasible but also potentially beneficial for users' psychological well-being. The developed system has shown its ability to generate images that align well with the detected emotions.

In conclusion, the results suggest that the ANIMO application can offer a new form of interaction between art and the user, where emotions play a central role in artistic creation. These findings suggest potential applications in art therapy and other contexts where emotional expression is crucial. Recommendations include further research on integrating more complex models to manage multiple emotions and exploring new techniques to enhance the fluidity and accuracy of emotion detection.

**Keywords** : Artificial Intelligence, Stable Diffusion, LoRa, Emotion Detection, Therapeutic Art, Web Application, Image Generation.

# ملخص

هذا التقرير هو ملخص للعمل الذي تم تنفيذه كجزء من مشروع نهاية الدراسة الذي تم تنفيذه في مختبر المعرفة والذكاء الاصطناعي الموزع، بهدف الحصول على درجة الماجستير في الذكاء الاصطناعي والواقع الافتراضي. يقترح المشروع نهجًا مبتكرًا للعلاج بالفن، باستخدام الذكاء الاصطناعي التوليدي لتجاوز قيود الأساليب التقليدية، والتي غالبًا ما تعتبر غير كافية لنقل المشاعر الإنسانية بدقة. يعتمد المشروع الذي تم تطويره بالتعاون مع الفنانة الفرنسية لينا كيي، على نماذج محددة تتكيف مع كل عاطفة - الفرح والحزن والخوف والحياد والغضب والمفاجأة والاشمئزاز - لتوليد صور فنية بناءً على الحالة العاطفية المكتشفة لدى المستخدم.

وتتمثل المشكلة التي يتناولها هذا المشروع في فهم كيفية استخدام الذكاء الاصطناعي لخلق أعمال فنية مبنية على المشاعر الإنسانية، واستكشاف التأثير المحتمل لهذا التفاعل على رفاهية المستخدمين. ويتمثل الهدف الرئيسي في تطوير نظام قادر على اكتشاف المشاعر من تعابير الوجه في الوقت الحقيقي وتوليد صور تعكس هذه المشاعر، وبالتالي تقديم شكل من أشكال التعبير البصري المخصص. تستكشف أسئلة البحث كيف يمكن تطبيق نماذج الانتشار المستقر ونماذج LoRA لتكييف الصور المولدة مع المشاعر المكتشفة.

لتحقيق هذه الأهداف، تم تنفيذ نهج يعتمد على اكتشاف المشاعر من خلال كاشف معالم الوجه. ثم يتم استخدام المشاعر المكتشفة لتحديد نموذج LoRA المناسب، والذي يتم تحميله لضبط توليد الصور. تستخدم تقنيات توليد الصور تقنية الانتشار المستقر، جنبًا إلى جنب مع نماذج LoRA لتنقيح النتائج وفقًا للعواطف المكتشفة. كما قام المشروع أيضًا بتطبيق آليات إدارة التزامن لضمان سلاسة الكشف وتوليد الصور، وتجنب التضارب في العملية.

تُظهر النتائج الرئيسية لمشروع ANIMO أن استخدام الذكاء الاصطناعي في الإبداع الفني العاطفي ليس فقط ممكناً، بل من المحتمل أن يكون مفيداً للمستخدمين من حيث الرفاهية النفسية. وقد أثبت النظام الذي تم تطوير قدرته على توليد الصور التي يُنظر إليها على أنها تتماشى مع المشاعر المكتشفة.

وفي الختام، تُظهر النتائج التي تم الحصول عليها أن تطبيق ANIMO يمكن أن يقدم شكلاً جديداً من أشكال التفاعل بين الفن والمستخدم، حيث تلعب العواطف دوراً محورياً في الإبداع الفني. تشير هذه النتائج إلى تطبيقات محتملة في العلاج بالفن والسياقات الأخرى التي يكون فيها التعبير العاطفي أمرًا بالغ الأهمية. تشمل التوصيات إجراء المزيد من الأبحاث حول دمج نماذج أكثر تعقيدًا لإدارة المشاعر المتعددة واستكشاف تقنيات جديدة لتحسين طلاقة ودقة اكتشاف المشاعر.

**الكلمات المفتاحية:** الذكاء الاصطناعي، الانتشار المستقر، لورا، الكشف عن المشاعر، الفن العلاجي، تطبيق الويب، توليد الصور.

# Sommaire

# Liste des figures

# Liste des abréviations

**CIAD :** Connaissance et Intelligence Artificielle Distribuée

**MHA :** Multi-Head Attention

**AI :** Artificial Intelligence

**GenAI:** Generative Artificial Intelligence

**GAN:** Generative Adversarial Network

**LDM:** latent diffusion Model

**SD:** Stable Diffusion

**LoRa:** Low-Rank Adaptation

**ResNET:** Residual Network

**U-NET:** Unet Convolutional Neural Network

**RNN:** Recurrent Neural Network

**MLP:** Multi-Layer Perceptron

**GPT :** Generative Pre-trained Transformer

**UI :** User Interface

**API :** Application Programming Interface

**FastAPI:** Fast Application Programming Interface.

# Introduction Générale

Dans un monde où les émotions façonnent nos expériences quotidiennes, la quête d'une représentation fidèle et nuancée de ces sentiments est devenue un enjeu majeur, particulièrement dans le domaine de l'art-thérapie. L'art-thérapie, une approche thérapeutique qui utilise le processus créatif pour favoriser l'expression et la guérison, se heurte souvent à des limitations lorsque les méthodes traditionnelles peinent à capturer la complexité des émotions humaines de manière précise. La capacité à traduire ces émotions en œuvres artistiques reste un défi important pour les praticiens de l'art-thérapie, qui cherchent constamment à améliorer la précision et la personnalisation des représentations émotionnelles.

L'intelligence artificielle (IA), avec ses avancées récentes, se présente comme une solution prometteuse pour surmonter ces défis. En particulier, l'IA générative, qui englobe des technologies telles que les modèles de diffusion stable et les techniques LoRa, offre de nouvelles possibilités pour créer des images artistiques basées sur les émotions détectées. Ces technologies permettent de générer des contenus visuels qui reflètent avec précision les états émotionnels des individus, en utilisant des algorithmes sophistiqués pour ajuster les images en fonction des données émotionnelles fournies.

Le cadre théorique de ce mémoire repose sur l'intersection de l'art-thérapie et de l'IA générative. En collaborant avec l'artiste française Lina KHEI, qui a créé un ensemble d'images pour représenter diverses émotions — joie, tristesse, peur, neutralité, colère, surprise, et dégoût — ce projet vise à explorer comment l'IA peut enrichir l'art-thérapie en produisant des œuvres visuelles personnalisées en réponse aux émotions détectées chez les utilisateurs.

La problématique centrale de cette étude est d'évaluer comment l'IA générative peut améliorer les méthodes traditionnelles d'art-thérapie en offrant une transcription plus précise des émotions humaines. Pour ce faire, la recherche se concentre sur l'utilisation de détecteurs de points de repère faciaux pour identifier les émotions et sur l'application des modèles LoRa pour adapter les images générées. La méthodologie implique une analyse approfondie des techniques de détection émotionnelle et de génération d'images, ainsi qu'une évaluation des résultats obtenus pour déterminer leur efficacité et leur impact.

Ce mémoire vise à démontrer la faisabilité et les bénéfices de l'intégration de l'IA générative dans ce domaine. Il a pour objectif de conduire un projet qui met en lumière comment ces technologies changent la face de l'art basé sur les émotions et ouvrent de nouvelles perspectives pour améliorer le bien-être des personnes par l'art.

Le plan de cette synthèse se structure comme suit : **le premier chapitre** introduit le contexte général du projet, présente le laboratoire d'accueil CIAD, le cadre du projet, la problématique, les objectifs ainsi que la méthodologie de travail, notamment la méthode Scrum employée. **Le second chapitre** présente le cadre théorique, avec un état de l'art sur l'IA générative, les GAN et les

modèles de diffusion. Enfin, **le troisième chapitre** aborde la conception et la mise en œuvre du projet ANIMO, décrivant le flux de travail, les outils utilisés, et les résultats obtenus avant de conclure sur les perspectives.

# Chapitre 1 : Présentation de l'organisme d'accueil

## 2.1 Introduction

Dans ce chapitre, nous proposons une présentation détaillée de l'organisme d'accueil, structurée en plusieurs sections. La première partie offre un aperçu global de l'entité, suivie par une présentation de ses principaux domaines d'application. Nous examinerons ensuite les projets et collaborations majeurs auxquels l'organisme contribue. Enfin, nous mettrons en lumière les chiffres clés illustrant ses performances et son rayonnement. Ce chapitre vise à fournir un cadre contextuel précis, essentiel à la compréhension de l'environnement dans lequel s'inscrit notre projet.

## 2.2 Présentation de laboratoire d'accueil

### 2.2.1 Présentation CIAD

Le laboratoire CIAD (Connaissance et Intelligence Artificielle Distribuées) est un laboratoire public de recherche sous la tutelle de l'Université de Bourgogne (UB) et de l'Université de Technologie de Belfort-Montbéliard (UTBM).
Concentré d'intelligence artificielle… et humaine, le CIAD regroupe des équipes de l'UTBM et de l'université de Bourgogne. Aperçu de la démarche et des activités de ce laboratoire de recherche en lien direct avec le monde industriel et les préoccupations sociétales.

Elle sait poser un diagnostic médical et sélectionner de la musique à la demande, elle est capable de créer une œuvre d'art aussi bien que de diriger une voiture… L'intelligence artificielle s'impose dans tous les domaines et semble ne connaître aucune limite. Qu'on la juge avec inquiétude ou enthousiasme, elle ne manque pas de fasciner. Née pour simuler l'intelligence de l'homme, elle rassemble les méthodes et les technologies qui permettront à des machines de se comporter comme lui, de faire preuve de raisonnement et de prise de décision, de témoigner d'une intelligence comparable. Le CIAD, pour Connaissance et Intelligence Artificielle Distribuées, est un laboratoire de recherche dédié à certaines des multiples facettes de l'intelligence artificielle (IA). Né avec l'année 2019, le CIAD conjugue des compétences d'équipes de l'UTBM et de l'université de Bourgogne (UB). Il ne lui a pas fallu plus de 6 mois pour obtenir le label Carnot, attribué aux structures de recherche publique dans le but de favoriser les partenariats avec l'entreprise. « Le CIAD fait preuve d'une politique de valorisation très forte et active auprès de l'industrie, explique Stéphane Galland, enseignant-

*Figure 1 : CIAD*

chercheur en informatique à l'UTBM et directeur adjoint du laboratoire. Le label Carnot est une reconnaissance de cette démarche en faveur de de la valorisation dans les entreprises et donne la possibilité de s'intégrer au sein d'un réseau économique et scientifique ».

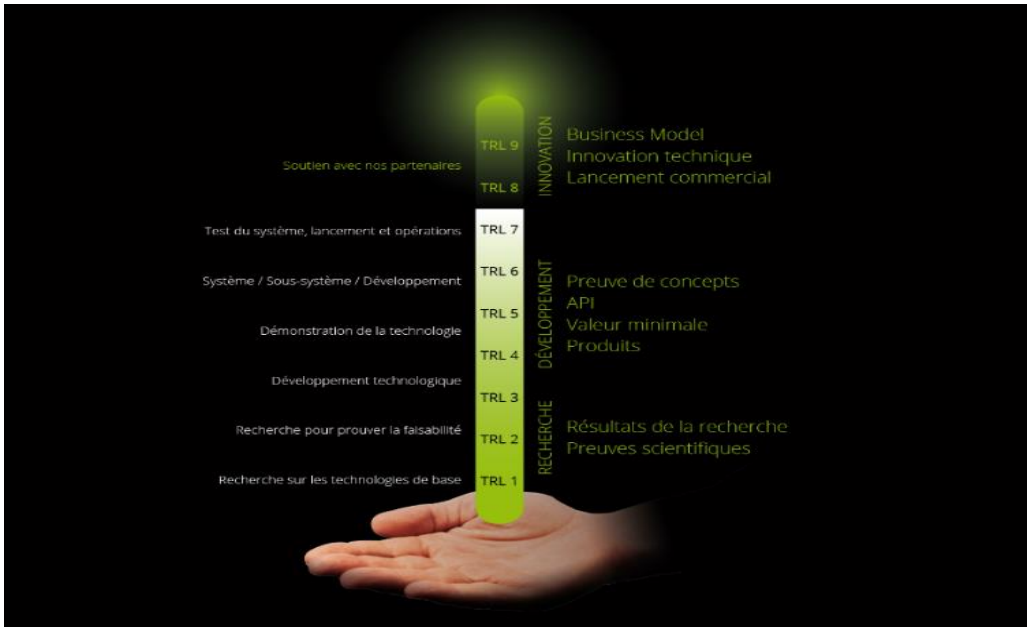



*Figure 2 : Processus de soutien à l'innovation*

## 2.3 Champs d'application

### 2.3.1 E-santé

- Développement de méthodes de perception active basées sur plusieurs capteurs pour le suivi humain et la reconnaissance d'objets, et la saisie d'objets inconnus destinés aux personnes handicapées et aux personnes âgées.
- Développement d'applications pour faciliter le diagnostic ophtalmologique.
- Analyse de sources de données de santé.

### 2.3.2 Smart City

- Amélioration de la sécurité des passages à niveau par l'intégration de systèmes permettant la gestion et la conception proactive de ses infrastructures.
- Construction d'un simulateur de train sur une plate-forme de simulation cyber-physique intégrant les composants réels des trains.
- Amélioration de la robustesse d'un système de perception d'objets pour la détection et le suivi dynamique dans des conditions défavorables (mauvais temps, trafic dense) pour des véhicules autonomes.
- Développement d'un outil logiciel pour l'assistance au diagnostic de pannes pour les locomotives.
- Construire un système de recommandation pour l'apprentissage en situation de mobilité.

- Développement d'un système de recommandation associant sémantique et métaheuristique pour résoudre des problèmes d'optimisation combinatoire liés à la combinaison pertinente d'offres touristiques selon un savoir-faire métier.

### 2.3.3 Industrie 4.0

- Amélioration de la vision informatique classique et les systèmes d'apprentissage en profondeur en prenant en compte les informations contextuelles de l'environnement et en effectuant un raisonnement en temps réel.
- Profilage dynamique du comportement des internautes dans un environnement e-marketing de Big Data.
- Elaboration d'un profilage dynamique d'internaute sur le web et recommandation publicitaire en temps réel.
- Développement d'une plateforme collaborative de partages de savoir-faire.

## 2.4 Projets et collaborations

Parmi les nombreux projets déjà réalisés ou en cours, on peut citer par exemple Explorys, né d'un partenariat entre l'UTBM et Alstom Transport : les connaissances liées aux composants mécaniques, électriques et électroniques d'un train sont modélisées, fournissent une aide à la maintenance par le biais d'outils interactifs et permettent la mise en œuvre de mécanismes de maintenance prédictive. Le projet 3L4AV, impliquant l'université technique tchèque de Prague, développe des méthodes d'apprentissage automatique en vue d'améliorer les systèmes de perception embarqués dans les véhicules autonomes : « Il s'agit d'assurer la détection et le suivi dynamique d'objets dans des conditions défavorables, comme le mauvais temps ou un trafic dense », précise Stéphane Galland. SAFER-LC a pour objectif d'assurer une meilleure sécurité des passages à niveaux qui sont des zones à haut risque pour la sécurité des automobilistes. Le projet prévoit la mise au point d'un système de vidéosurveillance avancé pour modéliser et analyser le comportement des utilisateurs.

Une vingtaine de projets collaboratifs, menés pour la plupart avec des partenaires industriels, sont inscrits à l'actif du CIAD. Politique d'utilisation des sols pour le développement de transports urbains, gestion de l'occupation de l'espace aérien des futures villes intelligentes par les drones, comportement des internautes dans un environnement e-marketing, reproduction des comportements de conduite en cas de brouillard dense…, tous développent et exploitent des méthodes à base d'intelligence artificielle, et notamment de systèmes multi-agent, pour aider à la mise au point de systèmes et d'outils adaptés aux futurs paysages économiques et sociaux. Le CIAD possède également une très forte expertise en modélisation d'environnement 3D, dans la conception et l'implantation de plateformes de jeux sérieux en réalité virtuelle, réalité augmentée ou sur dispositifs mobiles. Ces outils sont associés aux logiciels d'intelligence artificielle pour permettre une interaction plus aisée avec les utilisateurs.

## 2.5 Les chiffres clés de CIAD

Créé en janvier 2019, le CIAD regroupe une soixantaine de collaborateurs, dont de nombreux ingénieurs et une forte proportion de doctorants : ils ne sont pas moins de 28 à préparer leur thèse

sur un sujet en lien avec la modélisation numérique et les outils de l'intelligence artificielle. Les travaux de recherche développés au CIAD sont menés en lien direct avec les préoccupations de l'industrie : sur l'échelle TRL (*Technology Readiness Level*,) qui mesure le niveau de maturité technologique d'une recherche, les projets du CIAD atteignent le niveau 7, sur les 9 que comporte ce système d'évaluation international.


*Figure 3 : Chiffres clés de CIAD*

## 2.6 Conclusion

En conclusion, ce chapitre a permis de dresser un portrait clair et complet de l'organisme d'accueil, en mettant en lumière son environnement, ses principaux domaines d'application, ses projets et collaborations, ainsi que ses chiffres clés. Cette présentation a pour but de contextualiser le projet dans un cadre institutionnel solide, en soulignant l'importance et la pertinence des activités menées par l'organisme. L'analyse de ces éléments pose les fondations nécessaires à la compréhension du projet et à l'évaluation de son impact potentiel dans le cadre des travaux à venir.

# Chapitre 2 : Contexte général du projet

## 2.1 Introduction

Dans ce chapitre, nous exposons le cadre général de notre projet à travers deux sections distinctes. La première section aborde le cadrage du projet en mettant en lumière son contexte général, sa problématique et ses objectifs, offrant ainsi une vision claire de sa portée. La deuxième section détaille la conduite et le déroulement du projet, fournissant ainsi un aperçu des étapes clés et des processus mis en place. L'objectif de ce chapitre est de fournir une vue d'ensemble exhaustive du projet, établissant ainsi une base solide pour la suite de notre étude

## 2.2 Cadre général du projet

### 2.2.1 Contexte du projet

Dans le cadre de l'art-thérapie, un défi majeur est la capacité à capturer et représenter avec précision les émotions humaines. Les méthodes traditionnelles peuvent souvent se révéler insuffisantes pour traduire fidèlement la complexité des états émotionnels des individus. Le projet ANIMO se propose de surmonter ces limitations en développant une approche innovante, structurée en deux volets distincts mais complémentaires.
Le premier volet du projet ANIMO est consacré à la détection des émotions. Cette étape repose sur des techniques avancées pour analyser les expressions faciales des utilisateurs et identifier les émotions spécifiques qu'ils ressentent. L'objectif est d'obtenir une mesure précise et fiable des états émotionnels, permettant ainsi une compréhension plus profonde des ressentis individuels.

Le second volet se concentre sur la traduction des émotions détectées en œuvres artistiques visuelles. En utilisant les données émotionnelles recueillies, le projet vise à générer des images artistiques qui reflètent fidèlement les sentiments identifiés. Cette transformation des émotions en créations visuelles permet de créer des œuvres personnalisées, offrant une nouvelle dimension à l'expression artistique et enrichissant l'expérience de l'art-thérapie.
En résumé, le projet ANIMO aspire à améliorer les méthodes traditionnelles d'art-thérapie en offrant une approche intégrée qui allie détection précise des émotions et traduction innovante en œuvres artistiques. Cette démarche vise à fournir des représentations plus exactes et personnalisées des émotions humaines, enrichissant ainsi le domaine de l'art-thérapie.

### 2.2.2 Problématique

Dans le domaine de l'art-thérapie, les méthodes traditionnelles de traitement des émotions reposent souvent sur des approches basées sur l'interprétation humaine, telles que les dialogues verbaux ou les représentations artistiques non interactives. Ces méthodes, bien que valables, présentent des limites notables en termes de précision et de personnalisation. L'interprétation des émotions par les thérapeutes peut être influencée par des biais subjectifs, ce qui peut entraîner des erreurs et une

adéquation inexacte entre les émotions ressenties par les individus et leur représentation artistique. De plus, les approches actuelles manquent souvent de réactivité en temps réel, limitant la capacité à capturer et exprimer les émotions de manière dynamique et précise.

Parallèlement, dans le domaine de la création artistique numérique, les technologies disponibles jusqu'à présent, telles que les capteurs de mouvement ou de son, ne permettent pas une analyse fine et précise des émotions humaines. Ces technologies sont limitées par leur incapacité à capturer les nuances émotionnelles avec la profondeur nécessaire pour produire des œuvres artistiques véritablement réactives et personnalisées. Les œuvres interactives existantes peinent à intégrer les émotions des utilisateurs de manière significative, et la représentation visuelle des émotions reste souvent superficielle et déconnectée de l'expérience émotionnelle réelle des individus.

Face à ces défis, l'intégration de l'intelligence artificielle (IA) dans le domaine de l'art-thérapie représente une opportunité innovante pour surmonter ces limitations. En exploitant les capacités avancées de l'IA générative, telles que les modèles de diffusion stable et les techniques d'adaptation, il est possible de capturer les émotions humaines avec une précision accrue et de les traduire en œuvres artistiques de manière dynamique et personnalisée. Cette approche pourrait permettre une capture en temps réel des états émotionnels, offrant ainsi une réactivité immédiate et une personnalisation des créations artistiques en fonction des émotions détectées.

Le projet ANIMO s'inscrit dans cette dynamique en visant à développer un système intégré capable de détecter les émotions des utilisateurs avec une grande précision et de les convertir en représentations visuelles artistiques en temps réel. Cette démarche s'attaque à la problématique centrale de l'amélioration de la précision et de la personnalisation des représentations émotionnelles dans l'art-thérapie. En surmontant les limitations des méthodes traditionnelles et des technologies artistiques numériques existantes, ce projet aspire à enrichir l'expérience thérapeutique en offrant une approche plus précise, réactive et personnalisée pour la gestion des émotions humaines.

### 2.2.3 Objectifs

Le projet ANIMO vise à répondre aux limitations actuelles des méthodes d'art-thérapie et des technologies artistiques en développant une solution innovante qui intègre des avancées technologiques pour capturer, interpréter et traduire les émotions humaines en œuvres d'art. Les objectifs principaux du projet ANIMO sont les suivants :

**Capturer les Émotions avec Précision** : Développer une méthode pour capturer les émotions des utilisateurs de manière précise et en temps réel en utilisant des technologies avancées d'analyse faciale. Cette précision vise à surmonter les limitations des méthodes traditionnelles qui reposent sur des interprétations subjectives et potentielles erreurs humaines.

**Interpréter les Émotions en Temps Réel** : Concevoir un système capable de traiter et d'interpréter les données émotionnelles instantanément. Ce système doit être capable de transformer les émotions détectées en représentations visuelles qui reflètent fidèlement les états émotionnels des individus, offrant ainsi une réponse immédiate et précise.

**Créer des Œuvres Visuelles Artistiques** : Utiliser les émotions capturées pour générer des œuvres artistiques dynamiques et personnalisées. L'objectif est de produire des créations visuelles qui ne sont pas seulement réactives mais aussi esthétiquement enrichissantes, enrichissant ainsi

l'expérience artistique et thérapeutique.

**Offrir une Expérience Immersive** : Fournir une interaction immersive et intuitive qui permet aux utilisateurs de s'engager activement avec les œuvres générées. L'expérience doit être conçue pour engager les utilisateurs sur un plan émotionnel et esthétique, facilitant ainsi une meilleure compréhension et gestion de leurs états émotionnels.

**Améliorer l'Efficacité Thérapeutique** : Contribuer à l'amélioration des méthodes d'art-thérapie en offrant une approche plus objective et précise pour représenter les émotions. Ce projet vise à démontrer comment les technologies avancées peuvent enrichir les pratiques thérapeutiques en fournissant des outils pour une meilleure gestion des émotions et un soutien plus ciblé.

**Évaluer l'Impact et l'Engagement** : Mesurer l'impact des œuvres générées sur l'engagement des utilisateurs et leur perception des émotions. Évaluer comment la combinaison de la technologie et de l'art affecte l'interaction et la satisfaction des utilisateurs, ainsi que son potentiel pour améliorer les résultats thérapeutiques.

Ces objectifs visent à révolutionner la manière dont les émotions sont capturées, interprétées et exprimées dans le contexte de l'art et de la thérapie, en exploitant les capacités avancées de l'intelligence artificielle pour créer une expérience enrichissante et innovante.

## 2.3 Méthodologie de travail

### 2.3.1. Conduite du projet

La réussite de chaque projet nécessite une approche managériale structurée. Pour assurer le bon déroulement du projet ANIMO, nous avons opté pour la méthode Scrum d'une manière implicite, une approche agile de gestion de projet. Cette méthode nous permet de travailler de manière collaborative, en nous ajustant continuellement aux besoins du projet tout au long de son avancement.

Dans le cadre de ce projet, des réunions régulières sont organisées pour garantir une gestion efficace.

Chaque semaine, nous tenons des réunions avec mes deux encadrants pour présenter les progrès réalisés, discuter des problèmes rencontrés, et planifier les étapes suivantes. De plus, tous les quinze jours ou chaque mois, l'artiste Lina KHEI se joint aux réunions pour suivre l'avancement du projet, fournir des retours et proposer des idées supplémentaires. Ces réunions ont pour objectif de s'assurer que l'équipe reste alignée et que le projet avance de manière optimale, tout en permettant une adaptation rapide aux changements et aux nouvelles informations.

### 2.3.2 La méthode SCRUM

La méthode appliquée pour la gestion de ce projet est la méthode Scrum qui est une des méthodes de gestion de projet Agile. Son objectif est d'améliorer la productivité de l'équipe, tout en permettant une optimisation du produit grâce à des feedbacks réguliers avec les utilisateurs finaux. De plus, Scrum est une pratique agile élémentaire qui permet également une mise à l'échelle, autrement dit le déploiement progressif de l'agilité à l'échelle de l'entreprise. Scrum est de nos

jours l'approche agile la plus utilisée par les équipes de développeurs car elle promeut les valeurs du fameux Agile Manifesto : la collaboration avec le client, l'acceptation du changement, l'interaction avec les personnes et des logiciels opérationnels

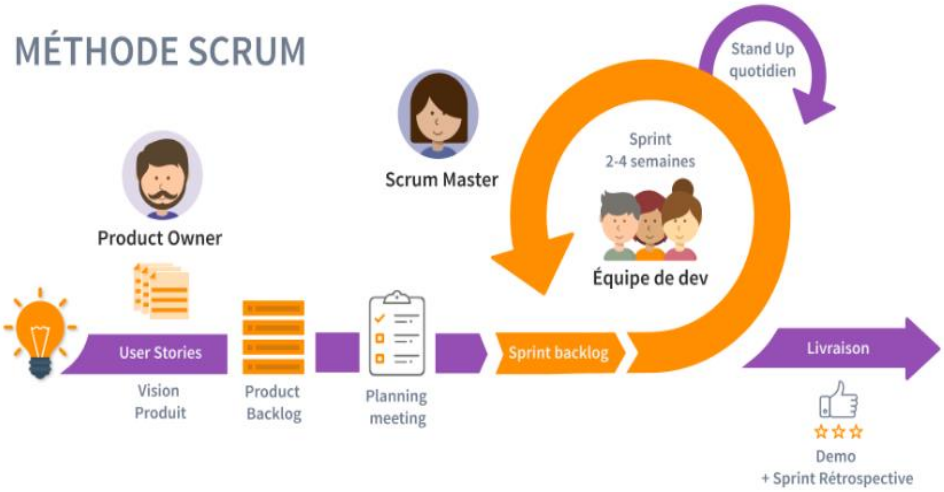



*Figure 4 : le processus de la méthode Scrum*

La figure précédente illustre le cycle da la méthode Scrum qui est au même le cycle qu'on a adopté lors du développement de notre solution.

#### 2.3.2.1 Les principales étapes de la méthode Scrum :

• **User Stories :** sont des descriptions courtes et simples des fonctionnalités ou des besoins du produit du point de vue de l'utilisateur final. Elles sont rédigées en langage naturel et se concentrent sur la valeur que ces fonctionnalités apportent à l'utilisateur.

*Application au projet ANIMO :*

**Exemple :** "En tant qu'utilisateur, je veux pouvoir voir des visuelles artistiques en temps réel basées sur mes émotions détectées afin de pouvoir mieux comprendre et exprimer mes sentiments "

**Objectif :** Capturer les besoins et les attentes des utilisateurs afin de guider le développement des fonctionnalités du projet

• **Product Backlog :** Est une liste ordonnée de tout ce qui est nécessaire dans le produit. Il contient toutes les user stories, fonctionnalités, améliorations, et corrections de bugs à réaliser. Le Product Owner est responsable de la gestion et de la priorisation de ce backlog.

**Contenu** : Le backlog pourrait inclure des éléments tels que la détection précise des émotions, la génération de visuels artistiques, l'intégration des technologies, et l'optimisation de la performance de l'application.

**Objectif** : Servir de liste de référence pour les tâches à réaliser tout au long du projet, permettant

de prioriser les éléments en fonction de leur importance et de leur valeur ajoutée.

**Sprint Planning Réunion :** La réunion de planification du sprint est une réunion où l'équipe Scrum se réunit pour définir les objectifs du sprint à venir et planifier les tâches à accomplir. Cette réunion se tient au début de chaque sprint (période de travail itérative de 2 à 4 semaines).

**Sprint Backlog :** Le sprint backlog est un sous-ensemble du product backlog qui contient les éléments sélectionnés pour le sprint en cours, ainsi que les tâches nécessaires pour accomplir ces éléments. C'est une liste de tâches spécifiques à réaliser pendant le sprint.

#### 2.3.2.2 Les rôles principaux de Scrum dans notre projet :

• **Scrum Master :** A pour objectif de faciliter l'organisation de l'équipe. Son rôle est de faire en sorte que celle-ci ne soit pas parasitée par des obstacles en s'assurant que la méthodologie Scrum est correctement appliquée, en organisant les réunions, en fixant les objectifs, et en facilitant la communication entre tous les membres de l'équipe pour garantir que le projet avance de manière fluide et cohérente, dans notre projet la Scrum Master est mon encadrant technique.

• **Product Owner :** Orienté vers le métier, il communique la vision du projet et les besoins du produit à l'équipe de développement. Dans le cadre de ce projet le Product Owner est l'artiste partenaire, elle est la principale partie prenante qui définit la vision artistique du projet et les besoins spécifiques en termes de fonctionnalités. Elle travaille en étroite collaboration avec l'équipe de développement, partage ses attentes, et s'assure que le produit final correspond à sa vision artistique et thérapeutique

• **Equipe de développement :** Responsable du développement technique du projet ANIMO, mon rôle est de mettre en œuvre les fonctionnalités prévues, de coder, tester et intégrer les éléments nécessaires pour réaliser le projet selon les directives données lors des réunions Scrum.

### 2.3.3. Planification du projet

La planification d'un projet est l'activité qui consiste à déterminer et à ordonnancer les tâches du projet, et à estimer leurs charges et déterminer les ressources nécessaires à leur réalisation.

Mais aussi à réfléchir à comment les risques projet seront gérés, à comment communiquer avec les parties prenantes du projet ...
Il est donc nécessaire de prendre le temps qu'il faut pour planifier un projet en détail avant de passer à l'action.

Et conformément à la méthode SCRUM, notre projet a été divisé en sprints et chaque sprint dure de 3 à 4 semaines. Pour rappel, un sprint est une itération de courte durée décomposant un processus de développement souvent complexe afin de le rendre plus simple et plus facile à réadapter et à améliorer en fonction du résultat des évaluations intermédiaires. Ci-dessous, le tableau 1 exposant quelques tâches que nous avons effectué lors de ce projet :

| | |
|---|---|
| Sprint 0: Initialisation et Préparation | - Obtention des accès nécessaires.<br><br>- Réunion initiale avec les encadrants pour définir les besoins généraux du projet.<br><br>- Suivi de formations en ligne sur les technologies nécessaires, comme défini par les encadrants. |
| Sprint 1: Collaboration avec l'artiste | - Réunion avec l'artiste partenaire pour comprendre ses besoins spécifiques.<br><br>- Présentation de l'atelier de l'artiste et de sa palette colorimétrique.<br><br>- Explication du besoin de détection des émotions de base pour guider la génération d'images.<br><br>- Analyse de la problématique à étudier et identification des limites de l'existant.<br><br>- Réception et organisation de la base de données préparée par l'artiste, triée par émotion. |
| Sprint 2: Conception et Installation | - Préparation de l'architecture globale du projet.<br><br>- Installation des frameworks nécessaires pour le développement.<br><br>- Recherche approfondie sur l'utilisation des GANs pour la génération d'images artistiques. |
| Sprint 3 : Développement et Test des Technologies | - Utilisation des GANs pour la génération des images artistiques.<br><br>- Étude des limites des GANs et exploration d'autres technologies d'IA générative.<br><br>- Mise en œuvre des modèles de diffusion stable et des modèles LoRA.<br><br>- Entraînement de chaque modèle par émotion. |
| Sprint 4 : Développement de la Détection des Émotions | - Développement d'un programme de détection des émotions en temps réel, en respectant la vie privée des utilisateurs (non-affichage du visage et aucune conservation des données). |

| | |
|---|---|
| Sprint 6 : Création de l'Interface et Intégration | - Création d'une interface graphique pour intégrer les systèmes de détection des émotions et de génération des images basées sur les émotions détectées. |

Tableau 1 : Taches réalisées au cours du projet ANIMO

## 2.4 Conclusion

Ce chapitre établit les bases sur lesquelles reposent les prochaines étapes du projet ANIMO. En détaillant le cadre général, la problématique et les objectifs du projet, ainsi que la méthodologie de travail, il souligne l'importance cruciale de l'intégration de la détection émotionnelle en temps réel et de la génération artistique. Ces éléments sont essentiels pour répondre aux besoins spécifiques du domaine de l'art-thérapie, en offrant une approche innovante qui vise à enrichir la précision et la personnalisation des expériences thérapeutiques. Les fondations posées dans ce chapitre orienteront le développement ultérieur du projet, en mettant en lumière les aspects clés à adresser pour atteindre les objectifs définis.

# Chapitre 3 : Cadre Théorique

## 3.1 Introduction

Dans ce chapitre, nous allons examiner les fondements théoriques qui sous-tendent le projet actuel, en mettant l'accent sur l'IA générative et la détection des émotions. Nous présenterons d'abord les travaux antérieurs relatifs à la génération d'images artistiques par IA puis à la détection des émotions, en mettant en évidence leurs limites et leur pertinence pour notre projet. Ensuite, nous établirons un lien entre ces travaux et notre projet afin d'arriver à une meilleure compréhension des enjeux théoriques. Enfin, nous pouvons exposer clairement le problème théorique que le projet ANIMO tente de résoudre et soumettre des hypothèses explicatives

## 3.2 Etat de l'art : IA Générative

### 3.2.1 Historique et Evolution

L'Intelligence Artificielle générative est un sous-domaine de l'intelligence artificielle qui se concentre sur la génération de contenus ou de solutions à partir d'un modèle appris à partir de données.
On peut alors dire que si l'Intelligence artificielle vise à imiter le cerveau humain, alors l'IA générative est la partie du cerveau qui permet de créer de nouveaux concepts en s'appuyant sur ce qu'il sait déjà. Plus concrètement, l'IA générative utilise des réseaux de neurones artificiels pour générer de nouvelles données comme des images, des sons ou même des textes de manière autonome. Cette technologie utilise des algorithmes de machine Learning tels que des modèles de diffusion appliqués à la restauration d'images, la segmentation et à la génération de contenu. Mais également les réseaux de neurones récurrents (RNN) plus approprié à la génération de textes ou plus récemment des Transformers.

L'histoire de l'intelligence artificielle générative (GenAI) est marquée par une série de développements technologiques et théoriques qui ont progressivement transformé la capacité la capacité à créer des contenus originaux. Voici un aperçu des étapes clés de son évolution :

**Années (1950s – 1980s)**

Les premières explorations de l'intelligence artificielle générative ont commencé dans les années 1950, avec des approches basées sur des algorithmes simples. Ces premiers systèmes utilisaient des règles grammaticales fixes ou des chaînes de Markov pour générer du contenu, souvent limité à des structures prédéfinies sans réelle créativité.

**L'Avènement des Réseaux de Neurones (1980s - 2000s)**

L'introduction des réseaux de neurones artificiels dans les années 1980 a marqué un tournant dans la génération de contenu. Les réseaux de neurones récurrents (RNN), conçus pour traiter des

séquences de données, ont été appliqués à la génération de texte. Ces modèles, bien qu'innovants, rencontraient des difficultés à maintenir la cohérence sur de longues séquences de données.

## L'Émergence des Autoencodeurs et des GANs (2010s)

Le début des années 2010 a vu l'émergence de deux innovations majeures : les autoencodeurs variationnels (VAE) et les réseaux antagonistes génératifs (GANs). Les VAE sont des modèles capables de générer de nouvelles données en apprenant des représentations compressées des données d'origine. Les GANs, introduits par Ian Goodfellow en 2014, se composent de deux réseaux de neurones distincts : un générateur, qui crée des données, et un discriminateur, qui évalue leur authenticité. Cette approche d'apprentissage par adversité a permis de produire des images et d'autres types de contenu avec un niveau de réalisme impressionnant.

## L'Ère des Transformateurs et des Modèles de Diffusion (2020s)

Les années 2020 ont été marquées par le développement des modèles de transformateurs, comme GPT (Generative Pre-trained Transformer) développé par OpenAI. Ces modèles utilisent des mécanismes d'attention pour traiter et générer des séquences de texte avec une cohérence et une pertinence contextuelle exceptionnelles. En parallèle, les modèles de diffusion, qui créent des données en simulant un processus inverse de diffusion, ont gagné en popularité pour la génération d'images. Ces modèles sont capables de produire des visuels détaillés et variés en apprenant à inverser la dégradation progressive des images.

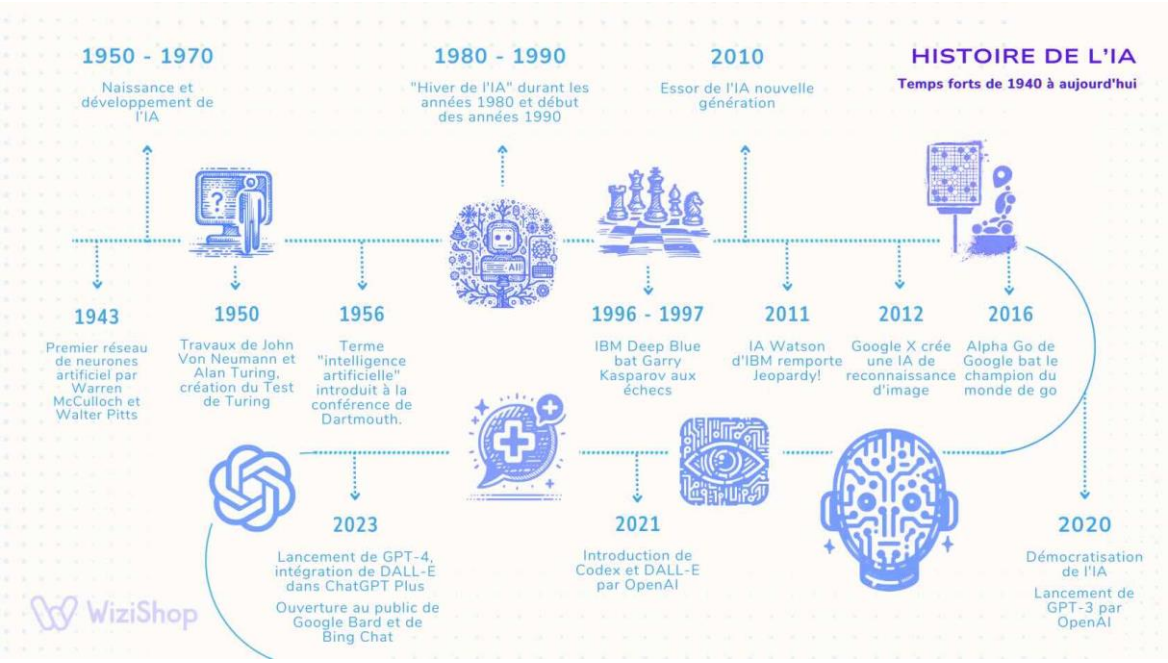



*Figure 5 : Histoire de l'Intelligence Artificielle*

### 3.2.2 Réseaux antagonistes génératifs (GAN)

Les réseaux antagonistes génératifs ou GANs (Generative Adversarial Network), sont des algorithmes d'apprentissage non supervisé à base de réseaux de neurones artificiels, qui permettent de modéliser et d'imiter n'importe quelle distribution de données. Ils peuvent être utilisés dans différents domaines (traitement d'images, de texte, de sons, …). Depuis leur invention en 2014, les GANs ont suscité un grand intérêt et plusieurs chercheurs ont souligné leur potentiel. Les GANs sont des modèles dits génératifs qui diffèrent des techniques traditionnelles d'analyse de données comme la classification. Alors que cette technique vise à apprendre à discriminer les données issues de différentes classes en fonction de leurs descripteurs, les algorithmes génératifs visent à faire le contraire : étant donnée une classe, les GANs cherchent à générer des données qui lui seraient associées.

Exemple : imaginons que l'on possède une base de données composée de peintures de différentes époques, parmi lesquels des tableaux de van Gogh. On pourrait entraîner un algorithme de classification pour déterminer si un nouveau tableau a été peint par Gogh ou non. On pourrait également entraîner un algorithme génératif pour créer de nouveaux tableaux dans le style de Gogh.

Concrètement, l'architecture d'un GAN est composée de deux réseaux de neurones, mis en compétition (voir schéma ci-dessous) qui s'entrainent ensemble de façon compétitive. Le générateur crée des données synthétiques à partir de bruit aléatoire, tandis que le discriminateur essaie de les distinguer des données réelles, Ainsi dans notre exemple, le discriminateur essayerait de détecter si une image est une peinture réalisée par Van Gogh ou s'il s'agit d'une peinture produite par le générateur.

Au fil du temps, le générateur devient de plus en plus performant pour produire des images réalistes.

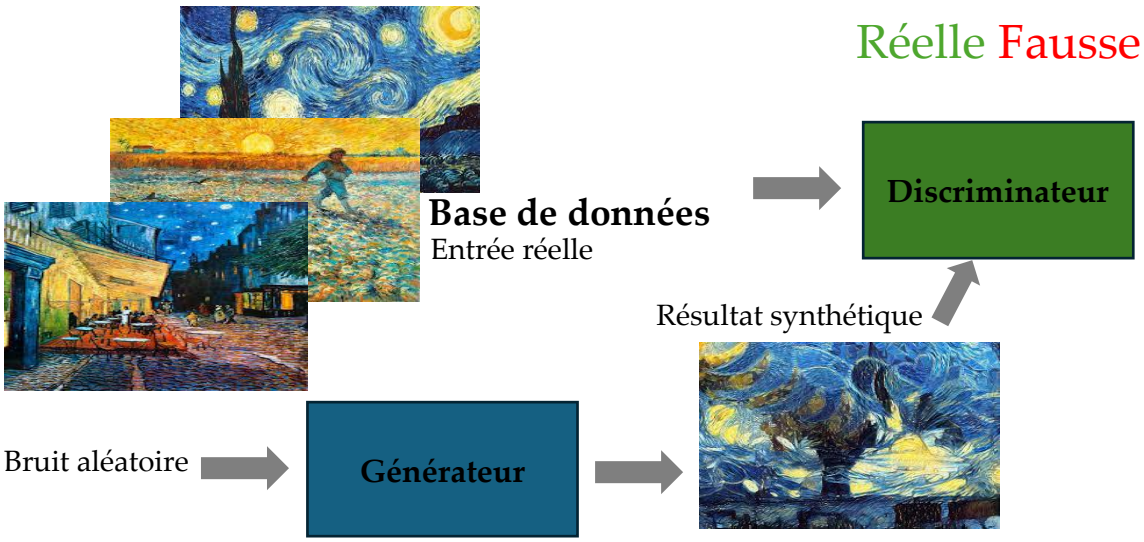



*Figure 6 : GAN- Fonctionnement du générateur et du discriminateur*

Les GAN fonctionnant de manière « antagoniste », formulé comme un défi d'apprentissage supervisé. Ici, le générateur aspire à créer des données si authentiques que le discriminateur est trompé environ 50 % du temps.
L'une des avancées notables dans le domaine des GAN est l'émergence des GAN à convolution profonde (DC-GAN), qui utilisent des couches convolutionnelles et se sont révélés particulièrement efficaces pour générer des images réalistes de haute qualité pour une myriade d'applications. Ces réseaux ont joué un rôle très important dans diverses réalisations en matière de génération d'images et de vidéos. Parmi les applications renommées, citons la transformation de l'esthétique des images via CycleGAN et la synthèse de visages humains hyperréalistes à l'aide de StyleGAN, comme l'illustre la plateforme « This Person Does Not Exist ».

Lors de l'entraînement, le GAN reçoit deux entrées, des données de bruit aléatoire et des données d'entrée qui ne sont pas étiquetées comme montré dans la figure précédente. À l'aide de ces deux entrées, il génère des données qui ressemblent aux données d'entrée. Étant donné que toutes les entrées du GAN ne sont pas étiquetées, le GAN est un type d'apprentissage automatique non supervisé.

En interne, GAN dispose de deux réseaux neuronaux qui sont en concurrence. Le but du réseau générateur est de tromper le réseau discriminateur et le but du réseau discriminateur est d'identifier correctement si l'entrée est réelle ou fausse.

#### 3.2.2.1 Générateur

Le rôle principal du générateur est de générer des données. Au départ, ces données sont susceptibles d'être constituées de bruit aléatoire, car le générateur démarre sans avoir beaucoup de connaissances sur la véritable distribution des données. Au fil du temps, à mesure que le GAN est formé, le générateur apprend à produire des données qui se rapprochent de la distribution réelle des données.

**Entrée :**

Le générateur prend en compte un vecteur de bruit aléatoire, généralement échantillonné à partir d'une distribution normale ou uniforme. Ce vecteur sert de graine ou de point de départ pour la génération de données.

**Architecture :**

Un élément clé responsable de la création de données actualisées et précises dans un réseau antagoniste génératif (GAN) est le modèle du générateur. Le générateur prend un bruit aléatoire en entrée et le convertit en échantillons de données complexes, tels que du texte ou des images. Il est généralement décrit comme un réseau neuronal profond.

La distribution sous-jacente des données d'apprentissage est capturée par des couches de paramètres pouvant être apprises dans sa conception grâce à l'apprentissage. Le générateur ajuste sa sortie pour produire des échantillons qui imitent étroitement les données réelles pendant l'apprentissage en utilisant la rétropropagation pour affiner ses paramètres.
La capacité du générateur à générer des échantillons variés et de haute qualité qui peuvent tromper

le discriminateur est ce qui fait son succès.

**Perte de générateur**

L'objectif du générateur dans un GAN est de produire des échantillons synthétiques suffisamment réalistes pour tromper le discriminateur. Le générateur y parvient en minimisant sa fonction de perte $J_G$. La perte est minimisée lorsque la probabilité logarithmique est maximisée, c'est-à-dire lorsque le discriminateur a une forte probabilité de classer les échantillons générés comme réels. L'équation suivante est donnée ci-dessous :

$$J_G = -\frac{1}{m}\sum_{i=1}^{m} \log D(G(z_i))$$

**Où,**

- $J_G$ Mesure dans quel point le générateur trompe le discriminateur.
- $\boldsymbol{\log D(G(z_i))}$ Représente la probabilité logarithmique que le discriminateur soit correct pour les échantillons générés.
- Le générateur vise à minimiser cette perte, en encourageant la production d'échantillons que le discriminateur classe comme réels.
  **($\boldsymbol{\log D(G(z_i))}$),** proche de 1

#### 3.2.2.2 Discriminateur

Un réseau neuronal artificiel appelé modèle discriminateur est utilisé dans les réseaux antagonistes génératifs (GAN) pour différencier les entrées générées des entrées réelles. En évaluant les échantillons d'entrée et en attribuant une probabilité d'authenticité, le discriminateur fonctionne comme un classificateur binaire.

Au fil du temps, le discriminateur apprend à différencier les données réelles issues de l'ensemble de données et les échantillons artificiels créés par le générateur. Cela lui permet d'affiner progressivement ses paramètres et d'augmenter son niveau de compétence.

Des couches convolutionnelles ou des structures pertinentes pour d'autres modalités sont généralement utilisées dans son architecture lorsqu'il s'agit de traiter des données d'image. Maximiser la capacité du discriminateur à identifier avec précision les échantillons générés comme frauduleux et les échantillons réels comme authentiques est l'objectif de la procédure d'apprentissage contradictoire. Le discriminateur devient de plus en plus discriminant en raison de l'interaction entre le générateur et le discriminateur, ce qui aide le GAN à produire des données synthétiques extrêmement réalistes dans leur ensemble.

**Perte de discriminateur**

Le discriminateur réduit la probabilité logarithmique négative de classer correctement les échantillons produits et réels. Cette perte incite le discriminateur à classer avec précision les échantillons générés comme des échantillons faux et réels avec l'équation suivante :

$$J_D = -\frac{1}{m}\sum_{i=1}^{m} \log D(x_i) - \frac{1}{m}\sum_{i=1}^{m} \log(1 - D(G(z_i))$$

- $\boldsymbol{J_D}$ Évalue la capacité du discriminateur à distinguer les échantillons produits des échantillons réels.
- La probabilité logarithmique que le discriminateur catégorise avec précision les données réelles est représentée par $\mathbf{log}\, \boldsymbol{D}(\boldsymbol{x_i})$.
- La probabilité logarithmique que le discriminateur catégorise correctement les échantillons générés comme faux est représentée par $\mathbf{log}(\mathbf{1} - \boldsymbol{D}(\boldsymbol{G}(\boldsymbol{z_i}))$ .
- Le discriminateur vise à réduire cette perte en identifiant avec précision les échantillons artificiels et réels.

**Perte MinMax**

Dans un réseau antagoniste génératif (GAN), la formule de perte minimax est fournie par :

$$\min_G \max_D (G, D) = [\mathrm{E}x\sim P_{data}[\log \mathrm{D(x)}] + \mathrm{E}z\sim P_{pz(z)}[log(1 - D(g(z)))]$$

**Où,**

- **G** est le réseau générateur et **D** est le réseau discriminateur.
- Les échantillons de données réelles obtenus à partir de la vraie distribution des données $P_{data}(x)$ sont représentés par x.
- Le bruit aléatoire échantillonné à partir d'une distribution antérieure $p_z(z)$(généralement une distribution normale ou uniforme) est représenté par z.
- D(x) représente la probabilité que le discriminateur identifie correctement les données réelles.
- D(G(z)) est la probabilité que le discriminateur identifie les données générées par le générateur comme étant authentiques.

### 3.2.2.3 Les étapes du fonctionnement d'un GAN

**Initialisation** : Deux réseaux neuronaux sont créés : un générateur (G) et un discriminateur (D).

- G est chargé de créer de nouvelles données, comme des images ou du texte, qui ressemblent étroitement à des données réelles.

- D joue le rôle de critique, en essayant de faire la distinction entre les données réelles (provenant d'un ensemble de données d'apprentissage) et les données générées par G.

**Premier mouvement du générateur :** G prend un vecteur de bruit aléatoire en entrée. Ce vecteur de bruit contient des valeurs aléatoires et sert de point de départ au processus de création de G. À l'aide de ses couches internes et des modèles appris, G transforme le vecteur de bruit en un nouvel échantillon de données, comme une image générée.

**Le tour du discriminateur** : D reçoit deux types d'entrées : des échantillons de données réelles provenant de l'ensemble de données d'apprentissage.
Les échantillons de données générés par G à l'étape précédente. La tâche de D consiste à analyser chaque entrée et à déterminer s'il s'agit de données réelles ou de quelque chose que G a inventé. Il délivre un score de probabilité compris entre 0 et 1. Un score de 1 indique que les données sont probablement réelles, et un score de 0 suggère qu'elles sont fausses.

**Le processus d'apprentissage :** C'est maintenant qu'intervient la partie contradictoire :
Si D identifie correctement des données réelles (score proche de 1) et des données générées (score proche de 0), G et D sont tous deux légèrement récompensés. Cela s'explique par le fait qu'ils font tous deux du bon travail.

Cependant, l'essentiel est de s'améliorer en permanence. Si D identifie toujours tout correctement, il n'apprendra pas grand-chose. L'objectif est donc que G finisse par tromper D.

**Amélioration du générateur :** Lorsque D qualifie par erreur la création de G de réelle (score proche de 1), c'est le signe que G est sur la bonne voie. Dans ce cas, G reçoit une mise à jour positive significative, tandis que D reçoit une pénalité pour s'être fait avoir.
Ce retour d'information aide G à améliorer son processus de génération afin de créer des données plus réalistes.

**Adaptation du discriminateur :** Inversement, si D identifie correctement les fausses données de G (score proche de 0), mais que G ne reçoit aucune récompense, D est encore renforcé dans ses capacités de discrimination.
Ce duel permanent entre G et D affine les deux réseaux au fil du temps.

Au fur et à mesure que l'entraînement progresse, G s'améliore en générant des données réalistes, de sorte que D a plus de mal à faire la différence. Idéalement, G devient tellement compétent que D ne peut plus distinguer de manière fiable les vraies données des fausses. À ce stade, G est considéré comme bien formé et peut être utilisé pour générer de nouveaux échantillons de données réalistes.

Au fur et à mesure que l'entraînement progresse, G s'améliore dans la production de données réalistes, ce qui rend plus difficile pour D de faire la différence. Idéalement, G devient tellement compétent que D ne peut plus distinguer de manière fiable les vraies données des fausses. À ce stade, G est considéré comme bien formé et peut être utilisé pour générer de nouveaux échantillons de données réalistes.

Les GAN (Generative Adversarial Networks) sont utilisés dans de nombreux domaines. Voici quelques-unes des utilisations largement reconnues des GAN :

#### 3.2.2.4 Les domaines d’utilisation des GAN

*Synthèse et génération d'images :* les GAN sont souvent utilisés pour des tâches de synthèse et de génération d'images. Ils peuvent créer des images fraîches et réalistes qui imitent les données d'entraînement en apprenant la distribution qui explique l'ensemble de données. Le développement d'avatars réalistes, de photographies haute résolution et d'œuvres d'art originales a été facilité par ces types de réseaux génératifs.

*Traduction d'image à image* : les GAN peuvent être utilisés pour des problèmes de traduction d'image à image, où l'objectif est de convertir une image d'entrée d'un domaine à un autre tout en conservant ses principales caractéristiques. Les GAN peuvent être utilisés, par exemple, pour faire passer des images du jour à la nuit, transformer des dessins en images réalistes ou modifier le style créatif d'une image.

*Synthèse texte-image :* les GAN ont été utilisés pour créer des images à partir de descriptions textuelles. Les GAN peuvent produire des images qui se traduisent par une description à partir d'un texte, tel qu'une phrase ou une légende.
Cette application pourrait avoir un impact sur le réalisme du matériel visuel produit à partir d'instructions textuelles.

*Augmentation des données :* les GAN peuvent augmenter les données existantes et accroître la robustesse et la généralisation des modèles d'apprentissage automatique en créant des échantillons de données synthétiques.
*Génération de données pour la formation :* les GAN peuvent améliorer la résolution et la qualité des images à faible résolution. En s'entraînant sur des paires d'images à basse et haute résolution, les GAN peuvent générer des images à haute résolution à partir d'entrées à basse résolution, ce qui permet d'améliorer la qualité des images dans diverses applications telles que l'imagerie médicale, l'imagerie satellitaire et l'amélioration des vidéos.

#### 3.2.2.5 Les avantages des GAN

***Génération de données synthétiques :*** Les GAN peuvent générer de nouvelles données synthétiques qui ressemblent à une distribution de données connue, ce qui peut être utile pour l'augmentation des données, la détection des anomalies ou les applications créatives.

***Résultats de haute qualité :*** Les GAN peuvent produire des résultats photoréalistes de haute qualité dans la synthèse d'images, la synthèse vidéo, la synthèse musicale et d'autres tâches.

***Apprentissage non supervisé :*** Les GAN peuvent être formés sans données étiquetées, ce qui les rend adaptés aux tâches d'apprentissage non supervisé, lorsque les données étiquetées sont rares ou difficiles à obtenir.

***Polyvalence :*** Les GAN peuvent être appliqués à un large éventail de tâches, notamment la synthèse d'images, la synthèse texte-image, la traduction image-image, la détection d'anomalies, l'augmentation des données, etc.

#### 3.2.2.6 Les limites des GAN

***Instabilité de la formation :*** Les GAN peuvent être difficiles à former, avec un risque d'instabilité, d'effondrement de mode ou d'échec de la convergence.

***Coût de calcul :*** Les GAN peuvent nécessiter beaucoup de ressources informatiques et être lents à entraîner, en particulier pour les images à haute résolution ou les grands ensembles de données.

***Surajustement :*** Les GAN peuvent sur ajuster les données d'apprentissage, produisant des données synthétiques trop similaires aux données d'apprentissage et manquant de diversité.

***Biais et équité*** *:* Les GAN peuvent refléter les biais et l'injustice présents dans les données d'apprentissage, ce qui conduit à des données synthétiques discriminatoires ou biaisées.

***Interprétabilité et responsabilité :*** Les GAN peuvent être opaques et difficiles à interpréter ou à expliquer, ce qui rend difficile de garantir la responsabilité, la transparence ou l'équité dans leurs applications.

### 3.2.3 Modèles de diffusion

Les modèles de diffusion sont des modèles génératifs qui apprennent à inverser un processus de diffusion pour générer des données. Le processus de diffusion consiste à ajouter progressivement du bruit aux données jusqu'à ce qu'elles deviennent du bruit pur. Grâce à ce processus, une distribution simple est transformée en une distribution de données complexe par une série de petites étapes incrémentales. Essentiellement, ces modèles fonctionnent comme un phénomène de diffusion inversée, où le bruit est introduit dans les données de manière directe et retiré de manière inversée pour générer de nouveaux échantillons de données. En apprenant à inverser ce processus, les modèles de diffusion partent du bruit et le débruitent progressivement pour produire des données qui ressemblent étroitement aux exemples d'apprentissage.

Principaux éléments des modèles de diffusion :

***Processus de diffusion directe :*** Ce processus consiste à ajouter du bruit aux données en une série de petites étapes. Chaque étape augmente légèrement le bruit, rendant les données progressivement plus aléatoires jusqu'à ce qu'elles ressemblent à du bruit pur.

***Processus de diffusion inverse :*** Le modèle apprend à inverser les étapes d'ajout de bruit. Partant d'un bruit pur, le modèle élimine le bruit de manière itérative, générant ainsi des données qui correspondent à la distribution d'apprentissage.

***Fonction de score :*** Cette fonction estime le gradient de la distribution des données concernant le bruit. Elle aide à guider le processus de diffusion inverse pour produire des échantillons réalistes.

#### 3.2.3.1 Processus de diffusion directe (Forward diffusion)

Dans ce processus, du bruit est ajouté progressivement aux données au cours d'une série d'étapes. Cela s'apparente à une chaîne de Markov où chaque étape dégrade légèrement les données en ajoutant du bruit gaussien.

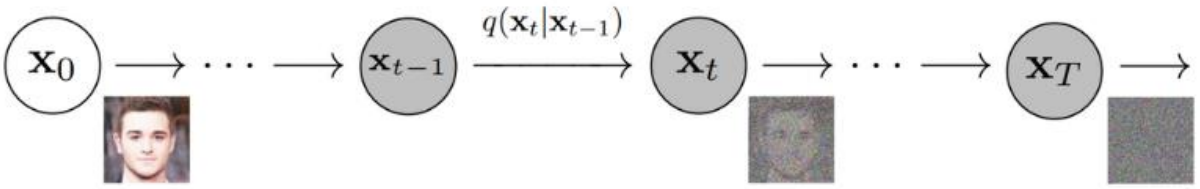



*Figure 7 : Le processus de diffusion directe (Forward diffusion)*

Mathématiquement, cela peut être représenté comme suit :

$$q(x_t|_1 x_t - 1) = N(x_t; \sqrt{\alpha_t} x_{t-1}, (1 - \alpha_t) I$$

**Où,**

- $x_t$ Est la donnée bruyante à l'étape t,
- $\alpha_t$ Contrôle la quantité de bruit ajoutée.

### 3.2.3.2 Processus de diffusion inverse (Reverse Diffusion) :

Le processus inverse vise à reconstruire les données originales en débruitant les données bruitées en une série d'étapes, en inversant la diffusion directe.

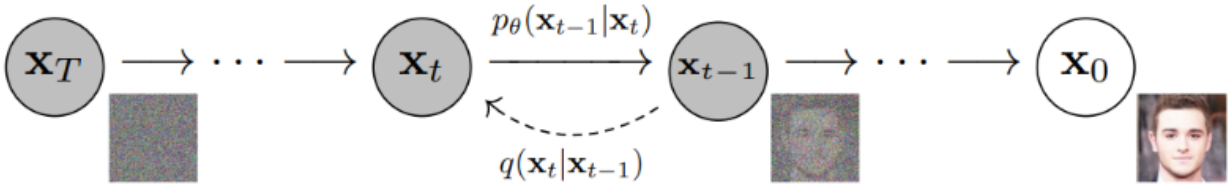



*Figure 8 : Processus de diffusion inverse*

Ce processus est généralement modélisé à l'aide d'un réseau neuronal qui prédit le bruit ajouté à chaque étape :

$$p_\theta(x_{t-1}|x_t) = N(x_{t-1}; \mu_\theta(x_t, t), \sigma_\theta(x_t, t))$$

**Où,**

- $\boldsymbol{\mu_\theta}$ Et $\boldsymbol{\sigma_\theta}$sont des paramètres appris.

### 3.2.3.3 Principe de fonctionnement des modèles de diffusion :

L'idée de base des modèles de diffusion est d'entraîner un réseau neuronal à inverser le processus de diffusion. Au cours de la formation, le modèle apprend à prédire le bruit ajouté à chaque étape du processus de diffusion. Pour ce faire, il minimise une fonction de perte qui mesure la différence entre le bruit prédit et le bruit réel.

**Processus directe (Diffusion) :**

Le processus direct consiste à corrompre progressivement les données $\boldsymbol{x_0}$ avec un bruit gaussien sur une séquence de pas de temps. Soit $\boldsymbol{x_t}$ représente les données bruitées au pas de temps t. Le processus est défini comme suit :

$$x_t = \sqrt{1 - \beta_t x_{t-1}} + \sqrt{\beta} t\, \epsilon$$

**Où,**

- $\boldsymbol{\beta_t}$ Est la programmation du bruit, un petit nombre positif qui contrôle la quantité de bruit ajoutée à chaque étape.
- $\in$ Est le bruit Gaussien.

Au fur et à mesure que t augmente, $\boldsymbol{x_t}$ devient de plus en plus bruyant jusqu'à ce qu'il se rapproche d'une distribution gaussienne.

### 3.2.3.4 Processus inverse (débruitage) :

Le processus inverse vise à reconstruire les données originales $\boldsymbol{x_0}$ à partir des données bruitées $\boldsymbol{x_T}$ au dernier pas de temps T. Ce processus est modélisé à l'aide d'un réseau neuronal pour approximer la probabilité conditionnelle $p_\theta(x_{t-1}, x_t)$

Le processus inverse peut être formulé comme suit :

$$x_{t-1} = \frac{1}{\sqrt{1 - \beta_t}} \left( x_t - \frac{\beta_t}{\sqrt{1 - \beta_t}} \in_\theta (x_t, t) \right)$$

**Où,**

- $\epsilon_\theta$ Est un réseau neuronal paramétré par θ qui prédit le bruit.

En général, le réseau neuronal est représenté de la manière suivante :

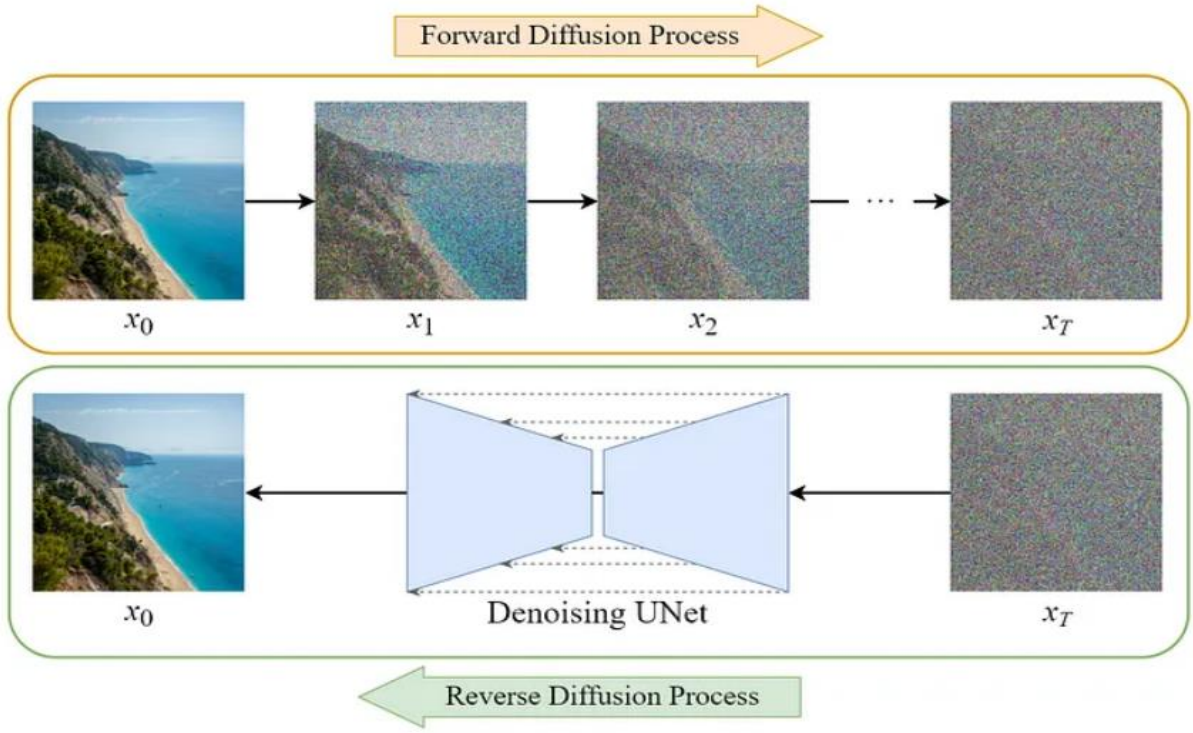



*Figure 9 : Vue d'ensemble du modèle de diffusion*

### 3.2.3.5 Entraînement des modèles de diffusion :

L'objectif d'apprentissage des modèles de diffusion consiste à minimiser la différence entre le bruit réel ϵ ajouté dans le processus direct et le bruit prédit par le réseau neuronal $\epsilon_\theta$

La fonction de score, qui estime le gradient de la distribution des données par rapport au bruit, joue un rôle crucial dans l'orientation du processus inverse. La fonction de perte est généralement l'erreur quadratique moyenne (EQM) entre ces deux quantités :

$$L(\theta) = E_{x_0,\epsilon,t}[||\epsilon - \epsilon_\theta(x_t, -t)||^2]$$

Alors à chaque époque :

- Un pas de temps aléatoire t sera sélectionné pour chaque échantillon d'apprentissage (image).
- Appliquer le bruit gaussien (correspondant à t) à chaque image.
- Convertir les pas de temps en encastrements (vecteurs).

Cela encourage le modèle à prédire avec précision le bruit et, par conséquent, à débruiter efficacement pendant le processus inverse.

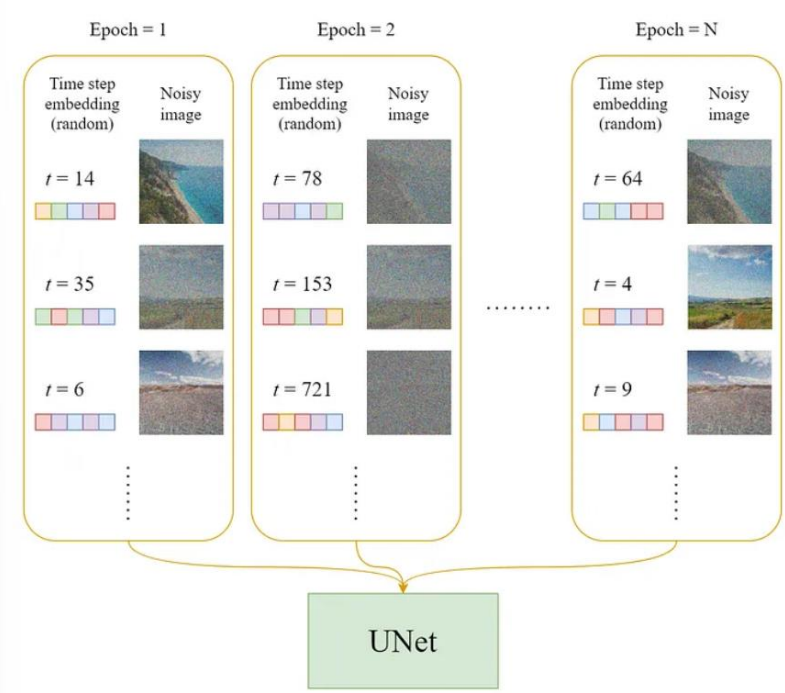



*Figure 10 : Ensemble de données pour l'apprentissage*

Pour chaque étape d'entrainement :

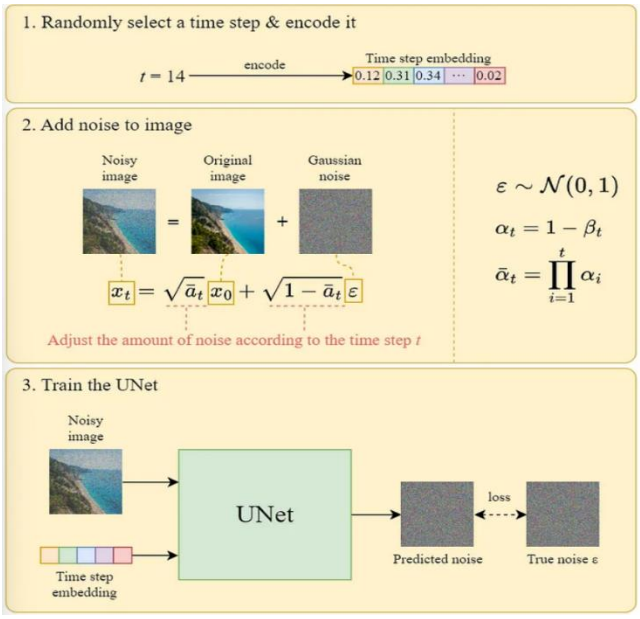



*Figure 11 : illustration de l'étape d'entrainement*

**Échantillonnage :**

L'échantillonnage consiste à peindre une image à partir d'un bruit gaussien. Le diagramme suivant montre comment nous pouvons utiliser le U-Net formé pour générer une image :

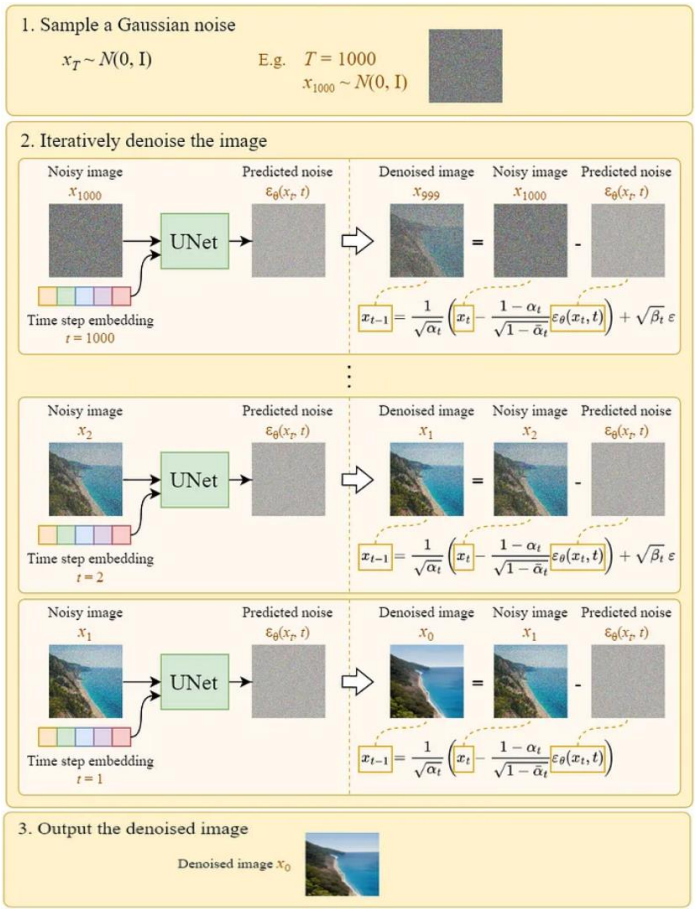



*Figure 12 : Illustration de l'échantillonnage*

### 3.2.3.6 Architecture :

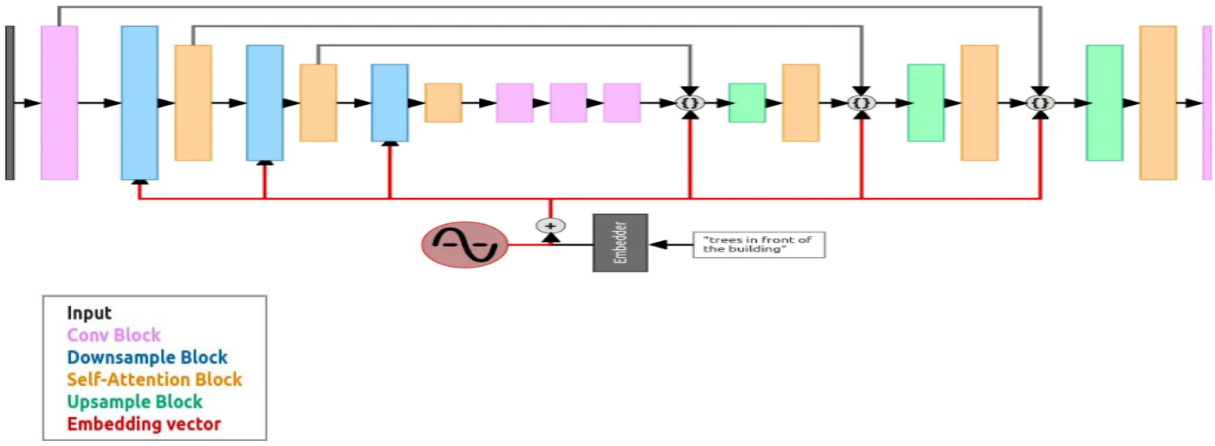



*Figure 13 : Architecture de model de diffusion*

L'architecture du modèle est une architecture U-Net modifiée. Elle est assez simple, mais elle se complique par la suite avec les améliorations apportées aux modèles de diffusion (par exemple, la diffusion stable a ajouté l'ensemble de la couche latente pour l'intégration des données d'image)

## Embeddings :

Avant d'aborder l'architecture des différents blocs, il convient d'examiner la manière dont les informations relatives au pas de temps et à l'invite sont transmises au réseau. La figure 5 donne l'impression que le modèle de diffusion se contente de traiter l'image d'entrée avec du bruit. Ce n'est pas vrai, chaque étape du processus ajoute une intégration avec des informations sur le pas de temps actuel et l'invite (si le modèle prend en charge l'invite, ce qui n'était pas le cas des premiers modèles de diffusion). Pour ce faire, nous devons utiliser un codage sinusoïdal pour coder le pas de temps t, et une sorte d'encodeur pour notre invite.

## Bloc ResNet :

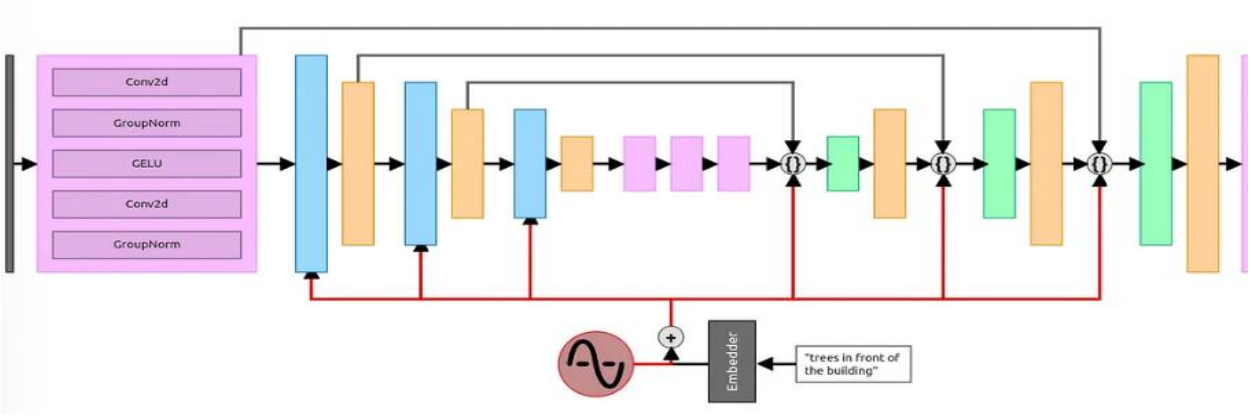



*Figure 14 : Bloc de ResNet*

Le premier bloc dont nous allons parler est le bloc ResNet.Dans cette version, le bloc ResNet est simple et linéaire. Ce bloc est utilisé ultérieurement dans le cadre des blocs de sous-échantillonnage et de suréchantillonnage.

## Bloc de sous-échantillonnage :

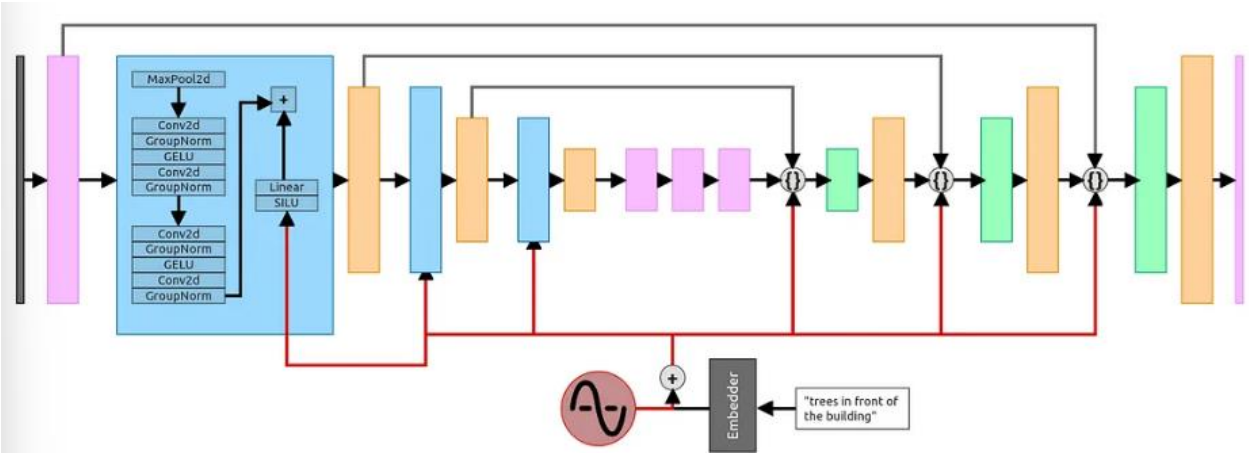



*Figure 15 : Bloc de sous-échantillonnage*

Le bloc de sous-échantillonnage est le premier bloc qui reçoit non seulement les données de la couche précédente, mais aussi les données relatives au pas de temps et à l'invite. Ce bloc a deux entrées et se comporte comme un sous-échantillonnage standard de l'architecture U-Net. Il reçoit une entrée et la réduit à la taille de la couche suivante. Il utilise la couche MaxPool2d (taille de noyau 2) qui divise par deux la taille de l'entrée (64x64 -> 32x32). Ensuite, nous avons 2 blocs ResNet (les mêmes que la couche entière juste avant dans la figure 7).

Les embeddings sont traités avec une unité linéaire sigmoïde et envoyés à travers une couche linéaire simple pour obtenir la même forme que la sortie du bloc ResNet. Ensuite, deux tenseurs sont ajoutés l'un à l'autre et envoyés au bloc suivant.

### Bloc d'auto-attention :

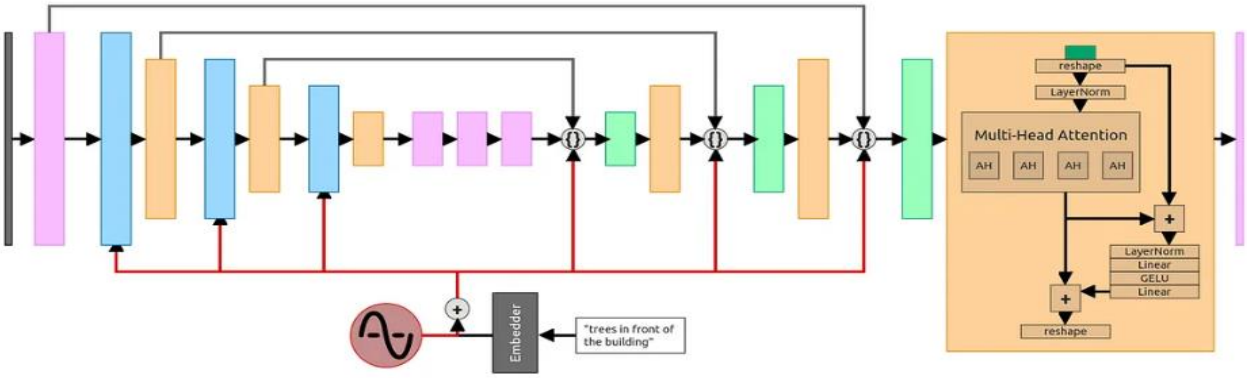



*Figure 16 : Bloc d'auto-attention*

Dans l'architecture U-Net modifiée, certains blocs ResNet ont été remplacés par des blocs d'auto-attention. Pour que ces blocs fonctionnent, il est nécessaire de remodeler les données d'entrée. Tous les blocs d'attention suivent une structure similaire, donc je vais décrire leur fonctionnement à travers le premier bloc, situé juste après le premier sous-échantillonnage. Ce bloc reçoit un tenseur sous-échantillonné de forme (128, 32, 32), et utilise une attention multi-têtes (Multi-Head Attention ou MHA) avec une dimension d'intégration fixée à 128 et 4 têtes d'attention. Bien que la dimension d'intégration varie entre les blocs en fonction de la taille de l'entrée, le nombre de têtes d'attention reste constant.

Pour appliquer la MHA, il est nécessaire de reformater l'entrée en créant les tenseurs Q(requête), K (clé) et V (valeur). L'entrée a une longueur de 128 et une taille spatiale de 32x32. Pour cela, les deux dernières dimensions sont aplaties et transposées, transformant le tenseur de (128, 32, 32) en (128, 1024), puis en (1024, 128). Ce tenseur est normalisé par une couche de normalisation avant d'être utilisé pour générer les trois tenseurs Q, K, et V.

À l'intérieur du bloc d'attention, deux connexions résiduelles sont ajoutées à la sortie du mécanisme d'auto-attention. La première connexion résiduelle combine l'entrée remodelée avec la sortie du bloc d'attention, et cette somme passe ensuite à travers une couche intermédiaire (normalisation -> linéaire -> GELU -> linéaire). La seconde connexion résiduelle ajoute la sortie de cette couche intermédiaire à la sortie du mécanisme d'attention. Enfin, le tenseur est remodelé pour revenir à sa forme initiale (1024, 128) -> (128, 32, 32), complétant ainsi le processus.

### Bloc de suréchantillonnage :

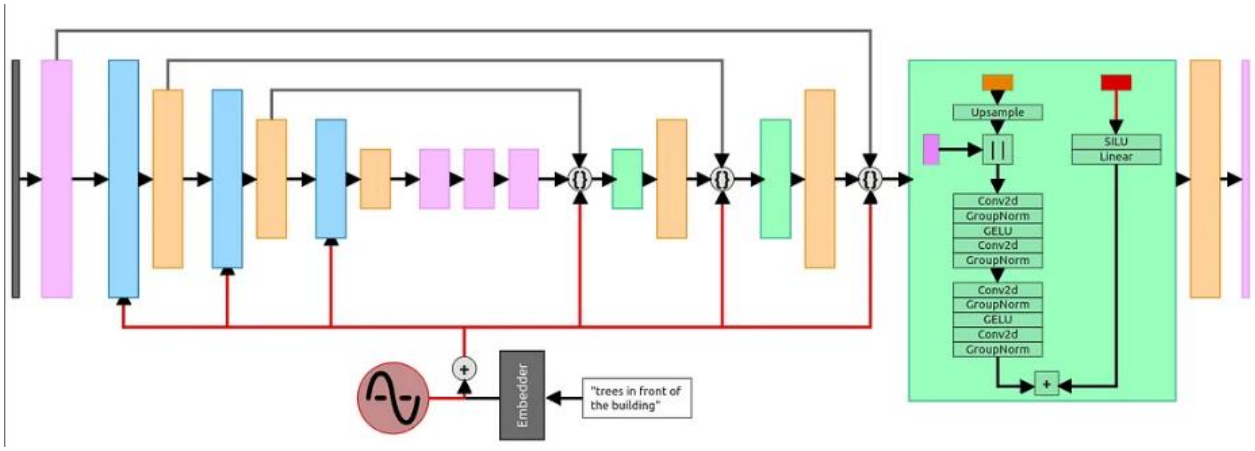



*Figure 17 : Bloc de suréchantillonnage*

Le suréchantillonnage est un peu plus compliqué car nous avons 3 entrées. L'entrée provenant de la couche précédente (dans le cas du 5ème bloc d'auto-attention) a des dimensions incompatibles. Comme il s'agit du bloc « upsample », il utilise une couche Upsample simple avec un facteur d'échelle de 2. Après avoir fait passer le tenseur d'entrée par la couche upsample, nous pouvons le concaténer avec la connexion résiduelle (du premier bloc ResNet). Ils ont maintenant tous deux la même forme (64, 64, 64). Les tenseurs concaténés sont ensuite envoyés à travers 2 blocs ResNet (comme nous l'avons fait dans le bloc de sous-échantillonnage).
La troisième entrée est (de la même manière que dans le bloc de déséchantillonnage) envoyée à travers SILU et la couche linéaire, puis ajoutée au résultat du deuxième bloc ResNet.
L'architecture entière se termine par la couche Conv2d, qui utilise un noyau de taille 1 pour ramener notre tenseur de (64,64,64) à (3, 64,64). Il s'agit de notre bruit prédit.

### 3.2.4 Diffusion Stable :

Le nom original de la diffusion stable est « Latent Diffusion Model » (LDM). Comme son nom l'indique, le processus de diffusion se déroule dans l'espace latent. C'est ce qui le rend plus rapide qu'un modèle de diffusion pur.

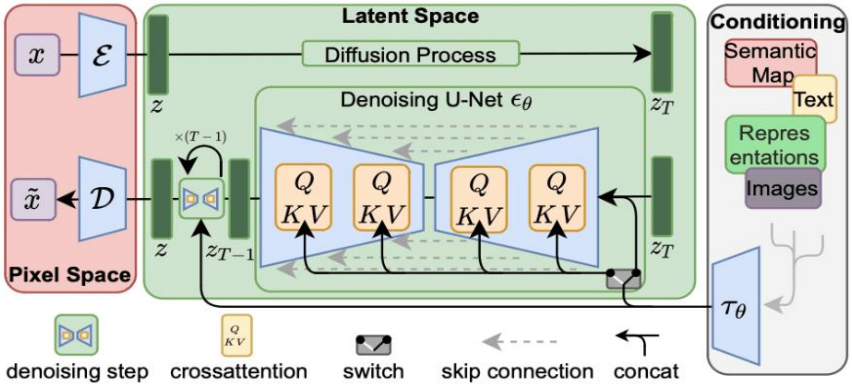



*Figure 18 : Schéma d'Architecture d'un Modèle de Diffusion Latente*

Départ de l'espace latent :

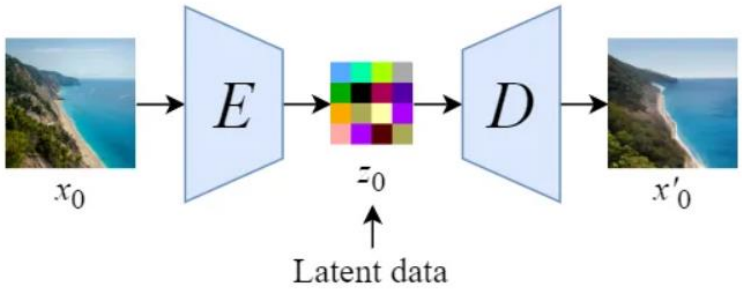



*Figure 19 : Auto-encodeur*

Nous commencerons par former un auto-encodeur pour qu'il apprenne à compresser les données de l'image en représentations de dimensions inférieures.

- En utilisant le codeur E entraîné, nous pouvons coder l'image de taille normale en données latentes de dimensions inférieures (données compressées).
- En utilisant le décodeur D entraîné, nous pouvons décoder les données latentes en une image.

## Diffusion latente :

Après avoir codé les images en données latentes, les processus de diffusion avant et arrière seront effectués dans l'espace latent.

- Processus de diffusion vers l'avant → ajouter du bruit aux données latentes.
- Processus de diffusion inverse → éliminer le bruit des données latentes.

## Conditionnement :

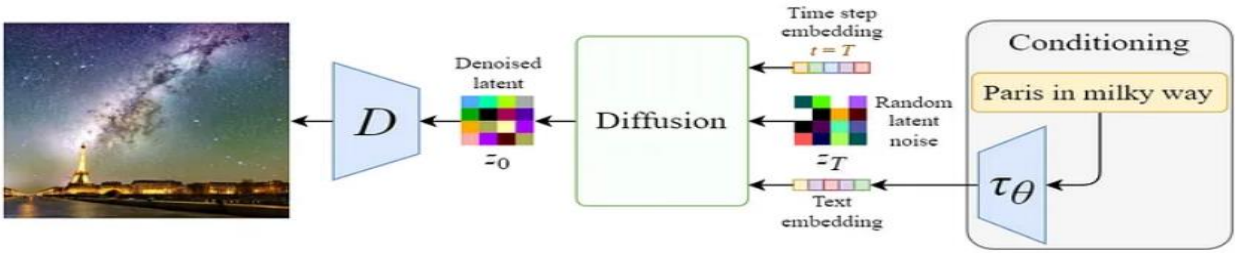



*Figure 20 : Présentation du mécanisme de conditionnement*

La véritable puissance du modèle de diffusion stable réside dans sa capacité à générer des images à partir d'invites textuelles. Pour ce faire, il suffit de modifier le modèle de diffusion interne afin qu'il accepte des entrées de conditionnement.

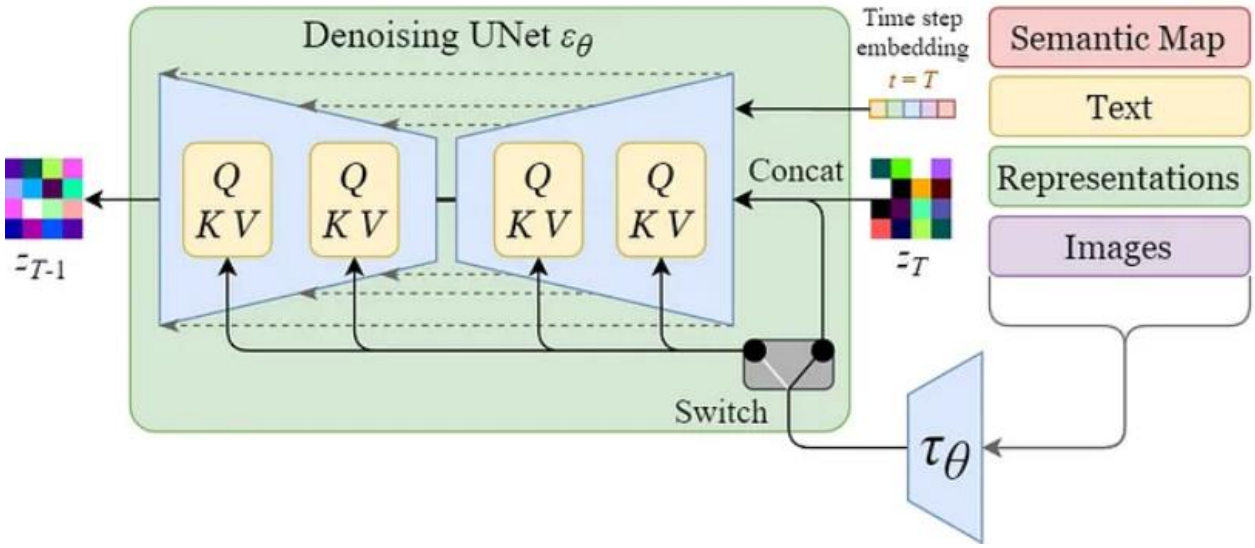



*Figure 21 : Présentation détaillée du mécanisme de conditionnement*

Le modèle de diffusion interne est transformé en un générateur d'images conditionnelles en augmentant son U-Net de débruitage avec le mécanisme d'attention croisée.
Le commutateur dans le diagramme ci-dessus est utilisé pour contrôler les différents types d'entrées de conditionnement :

- Pour les entrées textuelles, elles sont d'abord converties en embeddings (vecteurs) à l'aide d'un modèle linguistique $\tau\theta$ (par exemple, BERT, CLIP), puis elles sont mappées dans le U-Net via la couche (à têtes multiples) Attention (Q, K, V).

- Pour d'autres entrées alignées dans l'espace (par exemple, cartes sémantiques, images, inpainting), le conditionnement peut être effectué en utilisant la concaténation.

Entrainement :

$$z_0 = E(x_0)$$

$$z_t = \sqrt{\bar{\alpha}_t}\, z_0 + \sqrt{1 - \bar{\alpha}_t}\, \varepsilon$$

$$L_{\text{LDM}} = \mathbb{E}_{t, z_0, \varepsilon, y}\left[ \| \varepsilon - \varepsilon_\theta(z_t, t, \tau_\theta(y)) \|^2 \right]$$

Conditioning

*Figure 22 : Objectif d'entraînement pour le modèle de diffusion stable*

L'objectif de formation (fonction de perte) est assez similaire à celui du modèle de diffusion pure. Les seuls changements sont les suivants :

- Entrée des données latentes $z_t$ au lieu de l'image $x_t$.
- Ajout d'une entrée de conditionnement $\tau_\theta(y)$ au U-Net.

Échantillonnage :

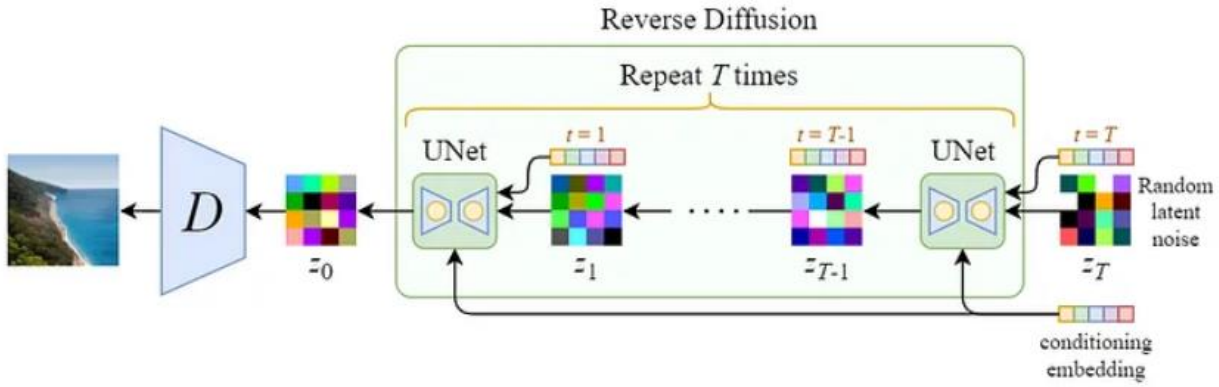



*Figure 23 : Processus d'échantillonnage par diffusion stable (débruitage)*

La taille des données latentes étant beaucoup plus petite que celle des images originales, le processus de débruitage sera beaucoup plus rapide.

Les principaux avantages sont la gratuité et l'utilisation intuitive de l'outil.

### 3.2.4.1 Les avantages de diffusion Stable :

- Simplicité d'utilisation
- Bonne résolution (pour la plupart des usages)
- Disponible gratuitement

### 3.2.4.2 Aperçu des inconvénients :

- Peut prendre du temps
- Sorties partiellement erronées
- Pour certains usages, la résolution n'est pas assez élevée
- Préoccupations juridiques
- Ne peut créer des images que sur des bases

### 3.2.4.3 Utilisation des LoRa dans diffusion stable :

Le modèle d'apprentissage profond de la diffusion stable est énorme. Le fichier de poids pèse plusieurs Go. Retraiter le modèle signifie mettre à jour un grand nombre de poids, ce qui représente beaucoup de travail. Parfois, nous devons modifier le modèle de Diffusion Stable, par exemple pour définir une nouvelle interprétation des invites ou pour que le modèle génère un style de peinture différent par défaut. En effet, il existe des moyens de faire une telle extension au modèle existant sans modifier les poids du modèle existant. Dans ce billet, nous découvrirons l'adaptation de rang faible LoRa (Low Rank Adaptation), qui est la technique la plus courante pour modifier le comportement de la diffusion stable.

LoRa, ou Low-Rank Adaptation, est une technique d'entraînement légère utilisée pour affiner les

modèles de diffusion stables et à grand langage sans avoir besoin d'un entraînement complet du modèle. L'ajustement complet de grands modèles (composés de milliards de paramètres) est intrinsèquement coûteux et prend beaucoup de temps. LoRa fonctionne en ajoutant un plus petit nombre de nouveaux poids au modèle pour l'entraînement, plutôt que de réentraîner l'ensemble de l'espace des paramètres du modèle. Cela réduit considérablement le nombre de paramètres pouvant être entraînés, ce qui permet des temps d'entraînement plus rapides et des tailles de fichiers plus faciles à gérer (généralement quelques centaines de mégaoctets). Les modèles LoRa sont donc plus faciles à stocker, à partager et à utiliser sur les GPU grand public.

En termes plus simples, LoRa revient à ajouter une petite équipe d'ouvriers spécialisés à une usine existante, plutôt que de construire une usine entièrement nouvelle à partir de zéro. Cela permet des ajustements plus efficaces et plus ciblés du modèle.

LoRa est une méthode de pointe proposée par les chercheurs de Microsoft pour adapter des modèles plus vastes à des concepts particuliers. Un réglage fin complet typique implique la mise à jour des poids de l'ensemble du modèle dans chaque couche dense du réseau neuronal. Aghajanyan et al. (2020) ont expliqué que les modèles sur-paramétrés pré-entraînés résident en fait dans une dimension intrinsèque faible. L'approche LoRa est basée sur cette constatation, en limitant les mises à jour des poids au résidu du modèle.

Supposons que $w_0 \in R^{d\times k}$ représente une matrice de poids pré-entraînée de taille $R^{d\times k}$(c'est-à-dire une matrice de d lignes et k colonnes en nombres réels), et qu'elle change par $\Delta w$
(La matrice de mise à jour), de sorte que les poids du modèle affiné sont les suivants :

$$w' = w_0 + \Delta w$$

LoRa utilise la technique qui permet d'abaisser le rang de cette matrice de mise à jour $\Delta w$ par décomposition des rangs, de telle sorte que :

$$\Delta w = B \times A$$

Ou B$\in R^{d\times r}$ et A$\in R^{r\times k}$, tel que r $\ll$ min(k,d).

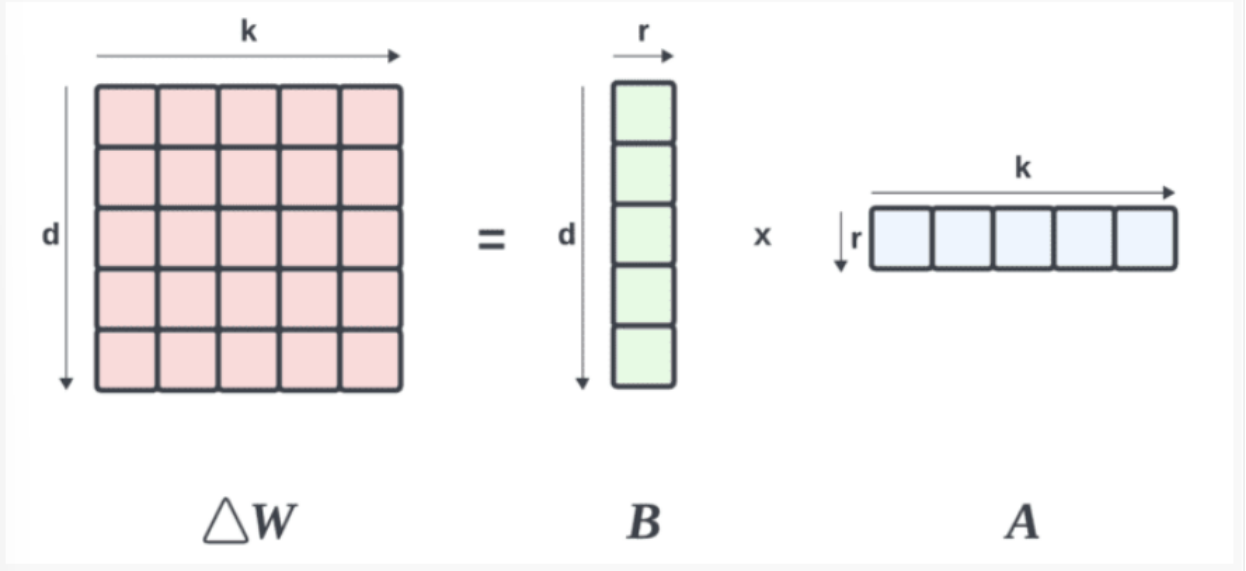



*Figure 24 : Décomposition d'une matrice en deux matrices de rang inférieur*

En gelant $w_0$(pour économiser de la mémoire), nous pouvons affiner A et B, qui contiennent les paramètres entraînables pour l'adaptation. Il en résulte que la passe avant du modèle affiné ressemble à ceci :

$$h = w'_x = w_0 x + BAx$$

Pour le réglage fin de la diffusion stable, il suffit d'appliquer la décomposition des rangs aux couches d'attention croisée (ombrées ci-dessous) qui sont responsables de l'intégration des informations de l'invite et de l'image. Plus précisément, les matrices de poids $W_o$, $W_Q$, $W_K$,et $W_V$ Dans ces couches sont décomposées pour abaisser le rang des mises à jour des poids. En gelant les autres modules MLP et en affinant uniquement les matrices décomposées A et B,

Les modules d'attention croisée (cross attention) peuvent être modifiés par LoRa.

### Exemples de modèles LoRa :

Il existe de nombreux modèles LoRa différents dans le contexte de la diffusion stable. Une façon de les classer est de se baser sur ce que fait le modèle LoRa :

***LoRa de personnage :*** ces modèles sont affinés pour capturer l'apparence, les proportions corporelles et les expressions de personnages spécifiques, souvent présents dans les dessins animés, les jeux vidéo ou d'autres formes de médias. Ils sont utiles pour la création d'œuvres d'art de fans, le développement de jeux et l'animation/illustration.

***Style LoRa :*** ces modèles sont affinés sur des œuvres d'art d'artistes ou de styles spécifiques afin de générer des images dans ce style. Ils sont souvent utilisés pour styliser une image de référence dans une esthétique particulière.

***LoRa vestimentaire :*** ces modèles sont affinés sur des œuvres d'art d'artistes ou de styles spécifiques afin de générer des images dans ce style. Ils sont souvent utilisés pour styliser une image de référence dans une esthétique particulière.

Voici quelques exemples :

*Figure 25 : Image créée avec le personnage LoRA « goku black [dragon ball super] »*


*Figure 26 : Image créée avec le style LoRA « Anime Lineart / Manga-like*

### 3.2.5 Comparaison des architectures :

Model de diffusion pur :

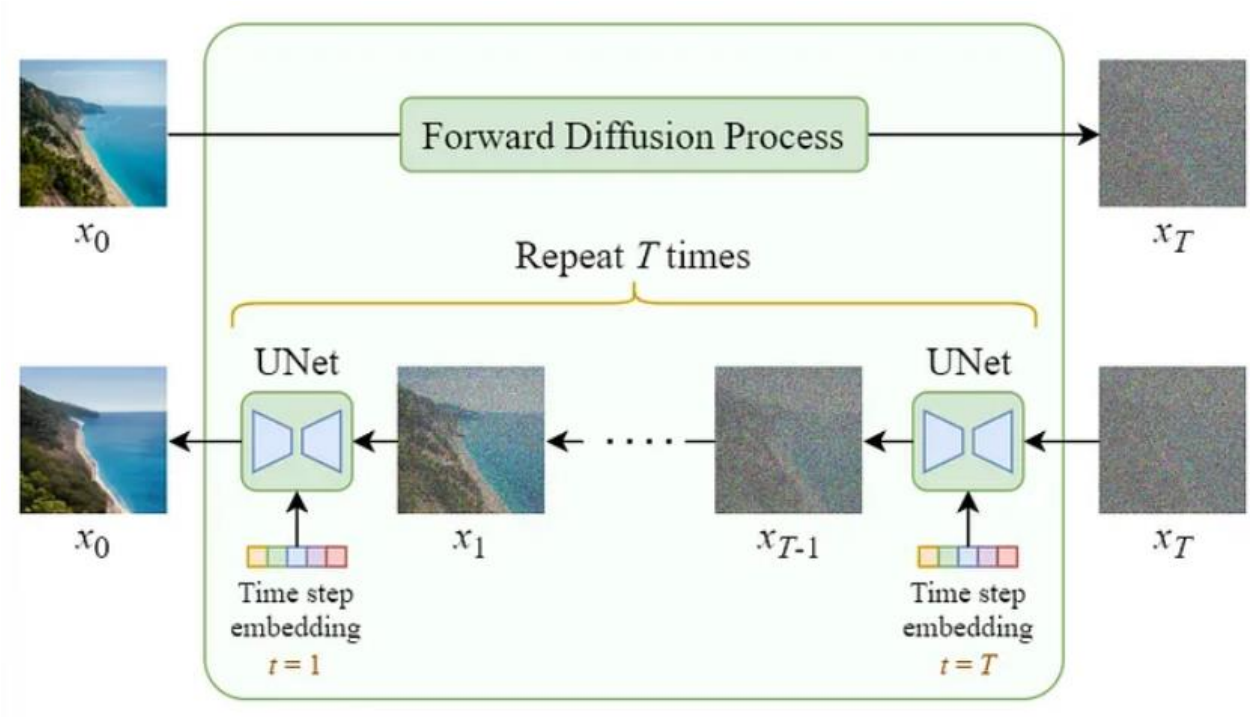



*Figure 27 : Diffusion stable (modèle de diffusion latente)*

## Diffusion stable (modèle de diffusion latente) :

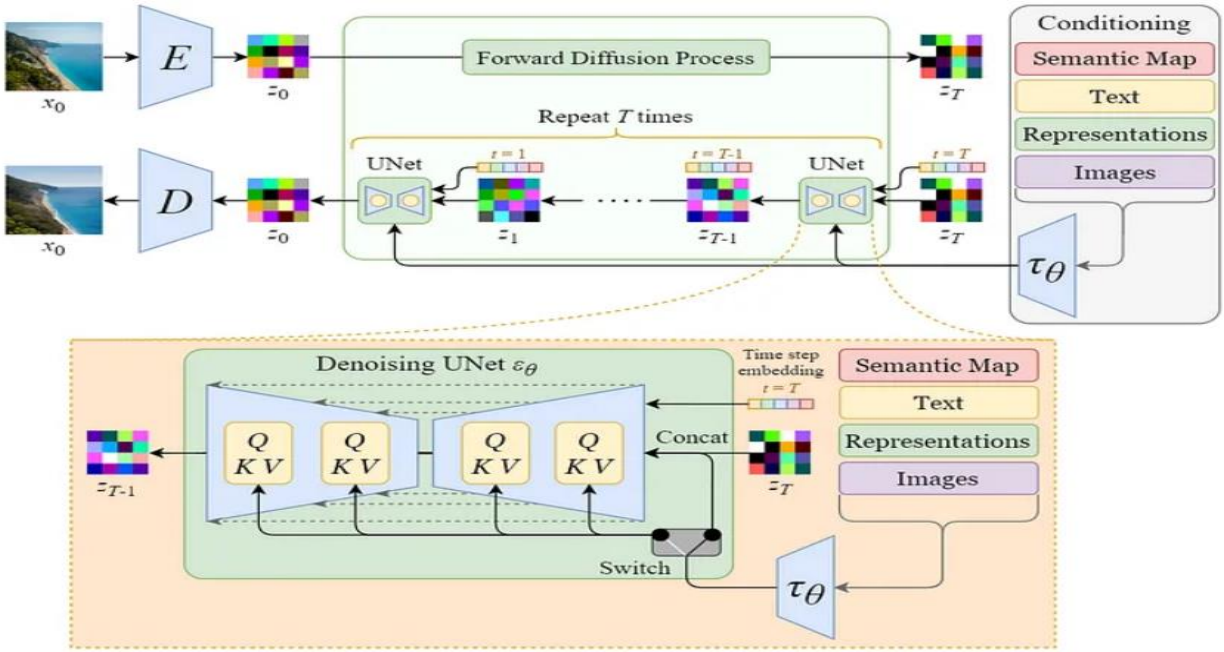



*Figure 28 : Architecture de stable diffusion*

En résumé, il existe plusieurs approches pour conditionner un modèle de diffusion sur la base d'un texte reçu en entrée. Il peut s'agir d'enchaîner des modèles de diffusion avec des modèles de langage intégrés ou de mettre en correspondance des données texte-image appariées et d'entraîner des modèles de diffusion sur ces données, ou encore d'utiliser un guidage sans classificateur avec

lequel nous pouvons ajuster le paramètre fidélité-créativité. L'objectif final de toutes ces approches est de s'assurer que l'image produite par le modèle est parfaitement alignée sur le contexte sémantique de l'invite textuelle donnée.

## 3.5 Conclusion

Ce chapitre a exploré l'évolution des approches en intelligence artificielle générative, en mettant en lumière les **GANs** et les **modèles de diffusion**, tels que **Stable Diffusion** et la technique **LoRA**. Les GANs, bien qu'innovants, présentent des limitations comme la difficulté d'entraînement et le mode collapse, tandis que les modèles de diffusion offrent une meilleure stabilité et qualité d'image, notamment avec le conditionnement dans l'espace latent. **Stable Diffusion** se distingue par son efficacité, et LoRA apporte une solution légère pour adapter les modèles. En conclusion, ces avancées théoriques, particulièrement dans la génération d'images artistiques, fournissent un cadre pertinent pour le projet **ANIMO**, visant à combiner génération et détection des émotions, en surmontant les défis rencontrés par les approches traditionnelles.

# Chapitre 4 : Conception et Mise en Œuvre

## 2.1 Introduction

Dans ce chapitre, nous allons détailler le processus de conception et de mise en œuvre du projet ANIMO. Qui consiste à générer des images artistiques en temps réel à partir des émotions détectées chez l'utilisateur. Après avoir étudié divers modèles génératifs tels que les GANs et les modèles de diffusion (comme Stable Diffusion), nous avons fait face à plusieurs défis liés à la qualité de l'image et aux performances en temps réel. Dans ce chapitre, nous allons expliquer comment nous avons surmonté ces obstacles en passant d'une approche basée sur les GANs à une implémentation avec Stable Diffusion v1.5.

## 4.2 Flux de travail du projet ANIMO :

Le projet ANIMO se compose de deux parties distinctes mais complémentaires, visant à repousser les limites de la transcription émotionnelle avec une grande précision La première présentée de manière générale pour préserver la confidentialité des détails, consiste en la conception d'une assise innovante intégrant des capteurs biométriques. Cette assise est conçue pour capturer les données émotionnelles de l'utilisateur, offrant ainsi une interface physique qui réagit dynamiquement à son état émotionnel.

Le second volet, sur lequel se concentre ce mémoire, se focalise sur la détection des émotions à partir de la webcam placée devant l'utilisateur et leur transcription en images artistiques. En associant des modèles spécifiques à chaque émotion — joie, tristesse, peur, neutralité, colère, surprise, et dégoût — le système vise à générer des œuvres visuelles personnalisées, avec une précision accrue, directement en fonction des émotions détectées en temps réel. L'objectif est de fournir une expérience immersive où l'exactitude de la transcription des émotions joue un rôle central, créant ainsi une nouvelle forme d'expression visuelle en art-thérapie, intimement liée aux états émotionnels de l'utilisateur.

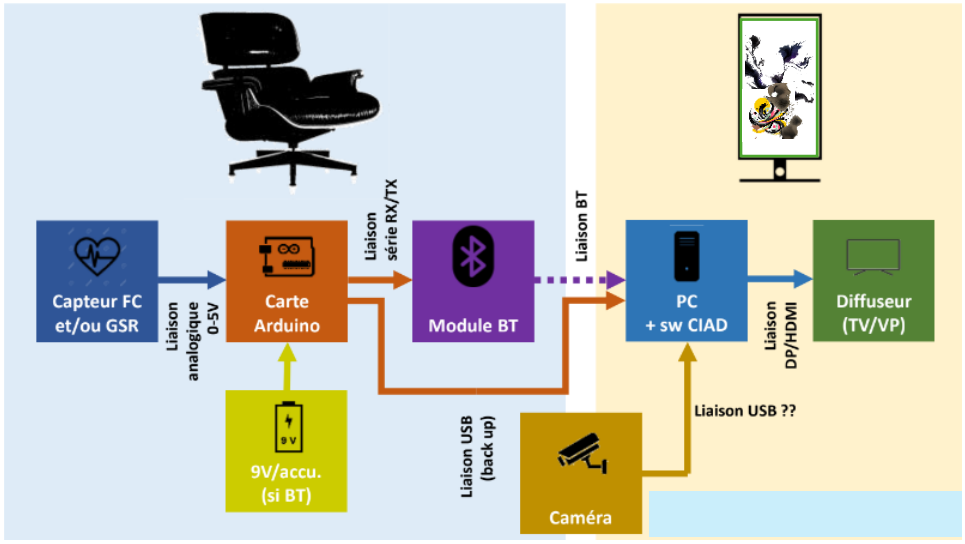



*Figure 29 : Le flux de travail général de ANIMO*

## 4.3 Conception Initiale : Utilisation des GANs

Au début du projet, nous avons exploré l'utilisation des Générative Adversial Networks (GANs) pour la génération d'images artistiques en fonction des émotions.Pour cette phase initiale, nous avons utilisé une base de données provenant de WikiArt en attendant la préparation de nos propres ensembles de données.Cette approche avait pour but de tester la viabilité des GANs pour notre application :

**Base de Données WikiArt :** La base de données WikiArt a été choisie en raison de sa vaste collection d'œuvres artistiques couvrant divers styles et genres. Cependant, plusieurs limitations ont été rencontrées :

- ***Résolution des Images :*** Les images disponibles dans la base de données WikiArt avaient une résolution variable, et la plupart n'étaient pas adaptées à notre besoin de générer des images en résolution 512x512 pixels. Cette limitation a entravé la qualité des images générées par les GANs, car une résolution plus basse entraîne généralement une perte de détails visuels importants.

- ***Diversité des Images*** : Bien que la base de données WikiArt offre une grande variété d'œuvres artistiques, elle présente certaines limites en termes de diversité émotionnelle. Les images disponibles n'étaient pas toujours représentatives des émotions spécifiques que nous cherchions à modéliser, ce qui a conduit à des résultats moins précis dans la génération d'images correspondant aux émotions étudiées.

| Style | Style Category | # Items Total | # Items Annotated |
|---|---|---|---|
| *Contemporary Art* | | | |
| | Minimalism | 2001 | 200 |
| *Modern Art* | | | |
| | Impressionism | 14862 | 200 |
| | Expressionism | 10629 | 200 |
| | Post-Impressionism | 7405 | 200 |
| | Surrealism | 6813 | 198 |
| | Abstract Expressionism | 4367 | 200 |
| | Cubism | 2963 | 200 |
| | Pop Art | 2004 | 200 |
| | Abstract Art | 1812 | 200 |
| | Art Informel | 1807 | 200 |
| | Color Field Painting | 1585 | 200 |
| | Neo-Expressionism | 1304 | 200 |
| | Magic Realism | 1289 | 153 |
| | Lyrical Abstraction | 1124 | 200 |
| *Post-Renaissance Art* | | | |
| | Realism | 13972 | 200 |
| | Romanticism | 10929 | 200 |
| | Baroque | 5498 | 200 |
| | Neoclassicism | 3450 | 197 |
| | Rococo | 2868 | 200 |
| *Renaissance Art* | | | |
| | Northern Renaissance | 2867 | 192 |
| | High Renaissance | 1465 | 104 |
| | Early Renaissance | 1405 | 119 |
| **Total** | | **151,151** | **4,105** |

Summary Table of the WikiArt emotion Dataset (Mohammad, Saif, and Svetlana Kiritchenko)

*Figure 30 : La base de données Wiki Art*

## Difficultés Rencontrées :

Malgré les avantages potentiels des GANs, les difficultés suivantes ont été observées :

- ***Qualité d'Image :*** Les images générées manquaient de cohérence et de fidélité par rapport aux émotions détectées, en partie en raison des limitations de résolution et de diversité des données d'entraînement.

- ***Complexité d'Entraînement :*** L'entraînement des GANs s'est révélé complexe, nécessitant des ressources computationnelles importantes et un ajustement minutieux des hyperparamètres pour obtenir des résultats satisfaisants.

*Figure 31 : Exemple des images générée à l'aide de GAN liées à l'émotion de la Joie*

En conséquence, ces défis ont conduit à une réévaluation de la méthode et à une transition vers l'utilisation de Stable Diffusion avec LoRA, une approche plus adaptée à nos besoins spécifiques en matière de génération d'images artistiques précises et diversifiées.

## 4.4 Présentation de l'Espace de Travail Kohya_ss et Stable Diffusion Automatic1111

Dans le cadre du projet ANIMO, l'entraînement des modèles LoRA (Low-Rank Adaptation) et leur validation sont des étapes cruciales pour générer des images artistiques en fonction des émotions détectées. Deux outils principaux sont utilisés dans cette démarche : Kohya_ss pour l'entraînement des modèles LoRA, et Stable Diffusion Automatic1111 pour tester et visualiser les résultats.
Voici une explication détaillée de l'espace de travail, englobant à la fois Kohya_ss pour l'entraînement des modèles et Stable Diffusion Automatic1111 pour le test et la visualisation.

### 4.4.1 Fonctionnement de Kohya_ss :

Kohya_ss est un ensemble d'outils puissants qui permet d'entraîner des modèles LoRA sur des modèles de génération d'images, comme Stable Diffusion, en utilisant des données spécifiques. Il est conçu pour effectuer un fine-tuning rapide et léger des modèles pré-entraînés sans nécessiter de grandes ressources matérielles ou des ensembles de données massifs.

Dans le cadre du projet ANIMO, Kohya_ss est utilisé pour entraîner des modèles LoRA à partir de données d'images associées à différentes émotions. Cela permet d'adapter un modèle Stable Diffusion pré-entraîné aux émotions détectées, afin de générer des images spécifiques reflétant ces émotions.

#### 4.4.1.1 Entrainement des modèles LoRa :

***Installation et configuration de kohya_ss :***

Pour former notre propre réseau LoRa, nous avons besoin d'un environnement de formation. Celui-ci est fourni par Kohya_ss de Bmaltais en open source et gratuitement.

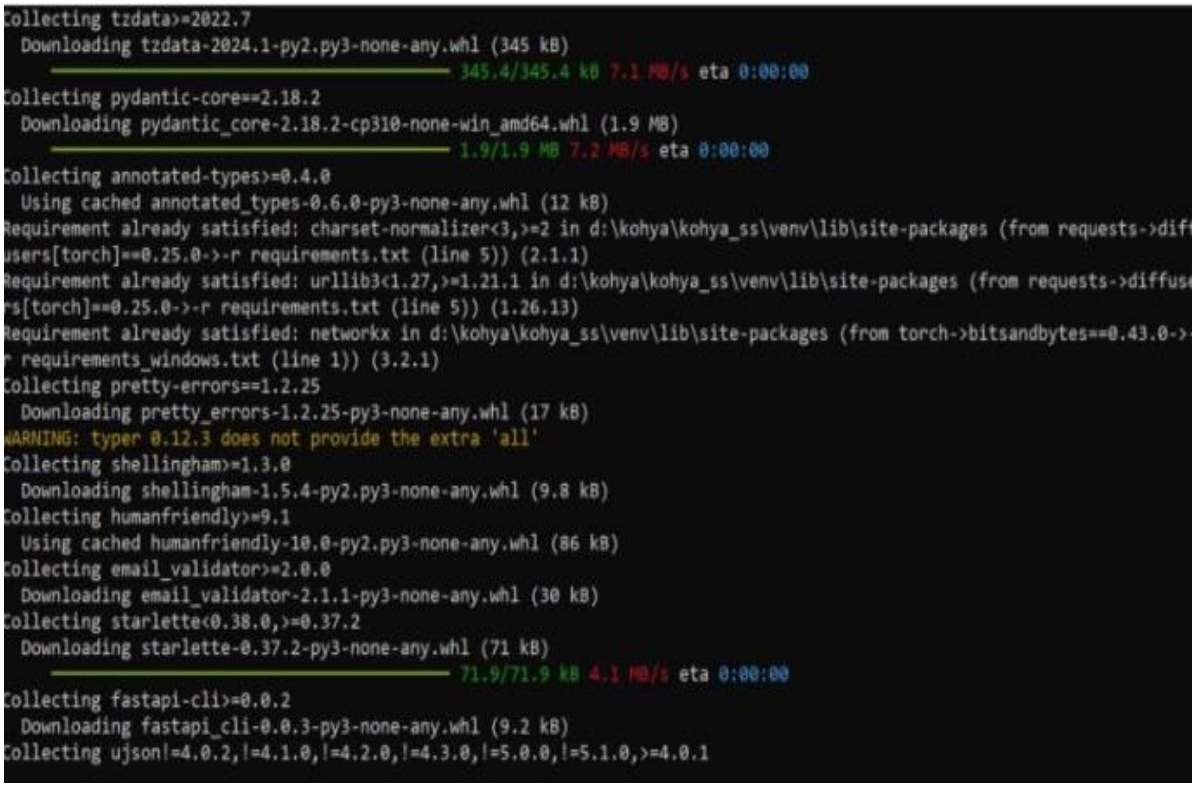



*Figure 32 : Installation de l'interface Kohya_ss*

Nous avons préparé les images pour être rassemblées et classées en fonction de l'émotion qu'elles représentent.
Ces images servent de base de l'entrainement du modèle LoRa. Par exemple, des images représentant la « colère » seront utilisées pour entrainer un modèle LoRa capable de générer des visuelles artistiques coléreuses.

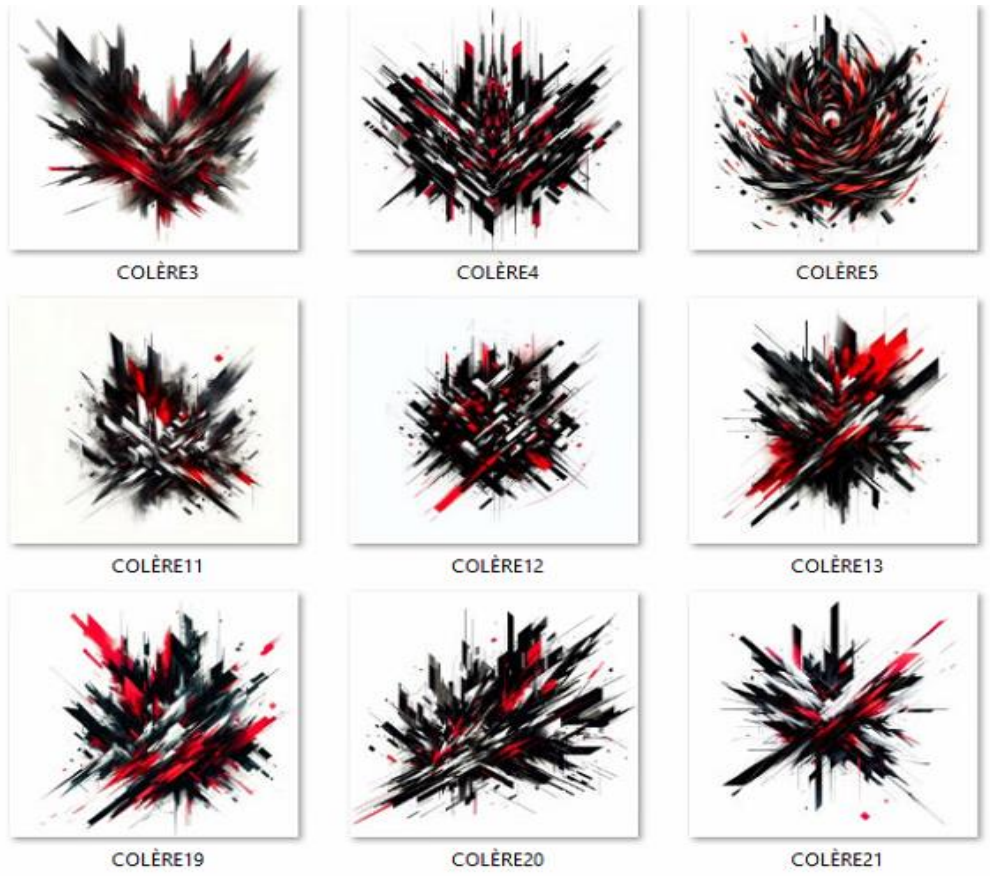



*Figure 33 : Extrait de l'ensemble de données représentant l'émotion "Colère"*

***Annotation des images :***

Avant de considérer nos images en tant que données d'entrainement, celles-ci doivent être annotées. Nous devons générer des descriptions de ce qui est montré sur l'image mais dans notre cas il suffit d'écrire l'émotion qu'elles représentent.

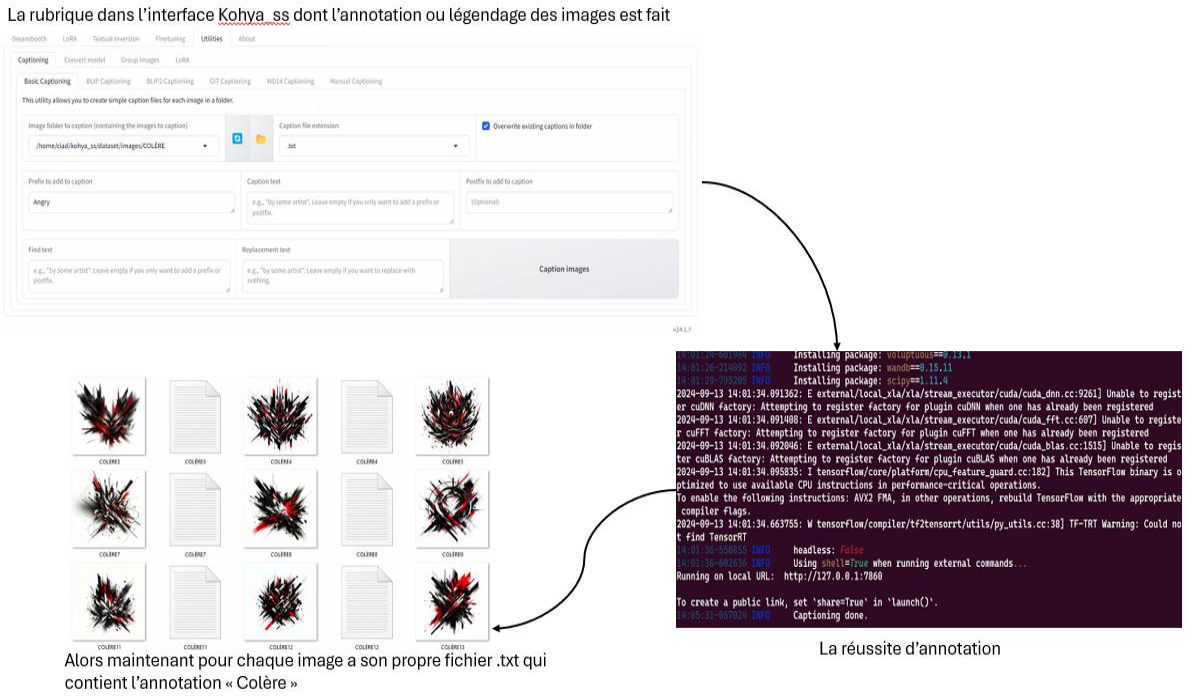



*Figure 34 : Le processus de légendage des images à l'aide de l'interface Kohya_ss*

## ***Préparation des données :***

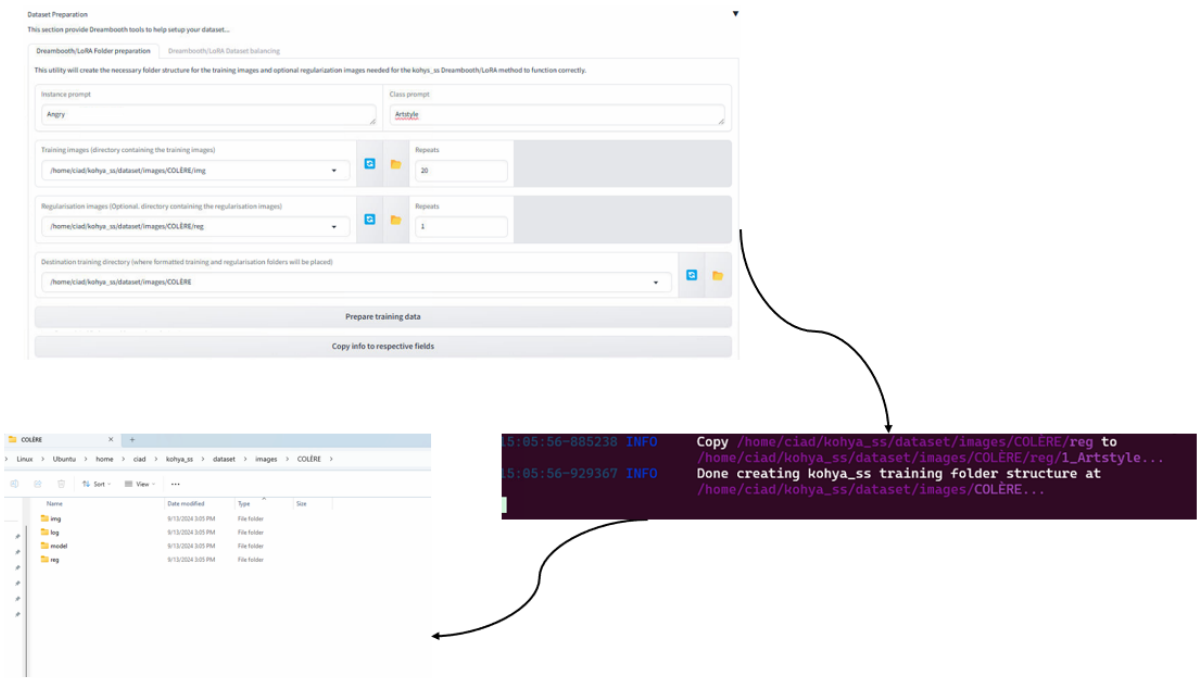
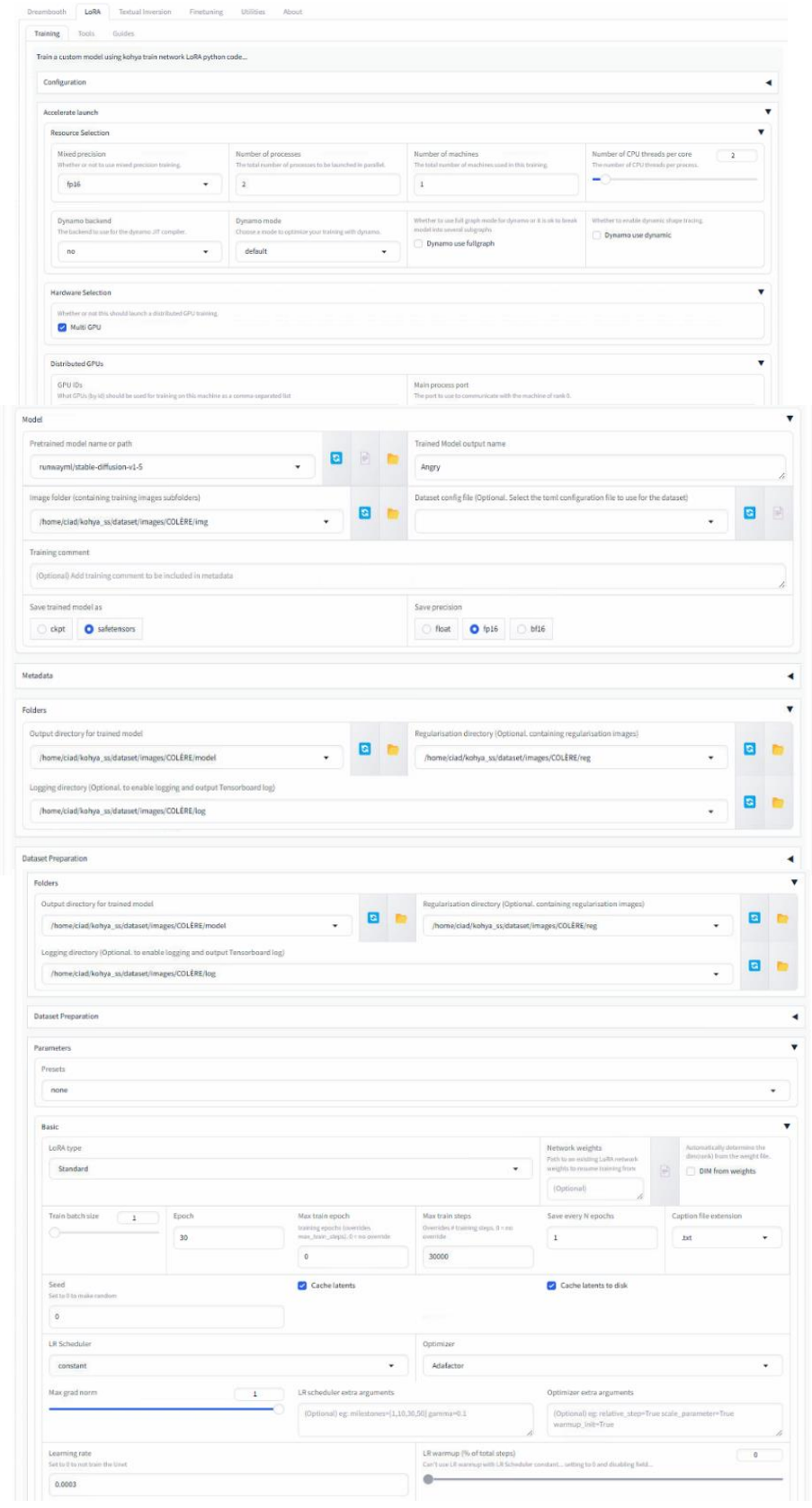


*Figure 35 : Processus de préparation des données*

Après que nous avons terminé la préparation de nos données avec succès comme montré dans la figure précédente nous pouvons maintenant commencer l'entrainement de notre LoRa.

***Entrainement du modèle :***



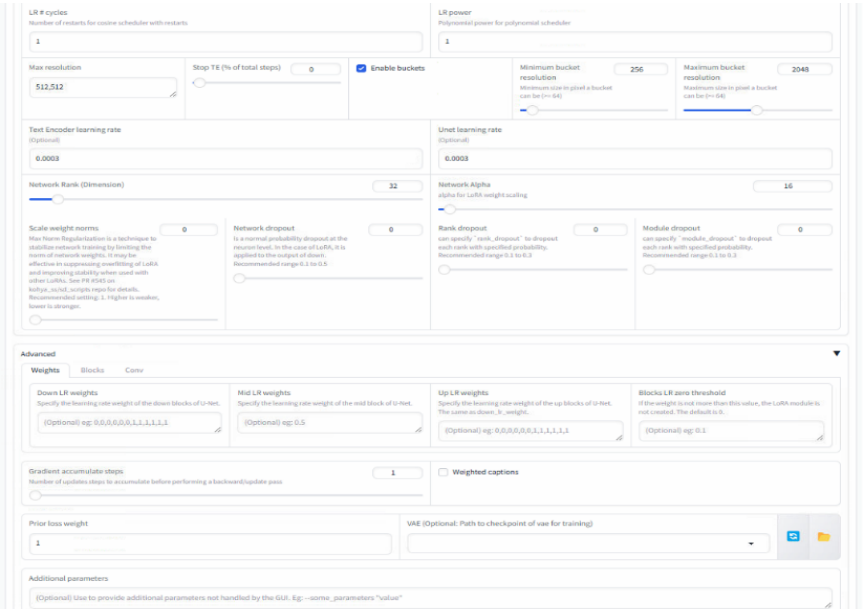


*Figure 36 : La préparation d'entrainement et choix de paramètres*

## Le choix des paramètres :

**LoRA Type – Standard**

- LoRA standard est utilisé pour ajuster certaines parties des poids d'un modèle pré-entraîné. Cela permet de former de nouveaux styles ou personnalités tout en préservant l'efficacité du modèle initial. Pour un style, cela permet de modifier la manière dont les images sont générées (couleurs, textures, formes), sans altérer le contenu général.
  C'est un moyen efficace d'adapter un modèle existant sans avoir à le réentraîner complètement.

**Train Batch Size**

- Le batch size représente le nombre d'images sur lesquelles le modèle est entraîné en parallèle. Pour capturer un style, une taille de lot plus élevée (5) peut aider à augmenter la variété et la généralisation du model, mais dans notre cas nous avons utilisé que 1 car on entraine un style artistique unique. Cela permet au modèle d'apprendre de manière plus fine et d'ajuster ses poids en fonction de chaque image.

**Epoch**

- Une epoch est une passe complète sur toutes les données d'entraînement. Répéter l'entraînement pendant plusieurs epochs (comme 10) permet au modèle de mieux apprendre et ajuster ses poids pour reproduire le style.
  En répétant l'entraînement plusieurs fois, le modèle affine ses ajustements, ce qui est important pour capturer des détails précis dans le style.

**Mixed Precision / Save Precision (BF16 ou FP16)**

- Le BF16 (ou FP16) permet d'utiliser moins de mémoire GPU tout en maintenant une précision suffisante pour les calculs d'entraînement.
- Utiliser une précision mixte réduit la consommation de ressources GPU, ce qui est crucial pour les grandes résolutions et les entraînements longs sur des styles complexes.

**Number of CPU Threads per Core**

- 2 threads par core signifient utiliser pleinement la puissance des processeurs modernes.

**Cache Latents and Cache Latents to Disk**

- Mettre en cache les latents signifie stocker des versions intermédiaires des données d'entraînement (les caractéristiques compressées des images) pour éviter de recalculer chaque fois.
  Cela réduit la charge de calcul et accélère le processus d'entraînement, surtout sur des styles qui nécessitent plusieurs passes d'images similaires.

**Learning Rate (0.001-0.004)**

- Le learning rate détermine la vitesse à laquelle le modèle ajuste ses poids pendant l'entraînement. Pour les styles, une plage de 0.001-0.004 est souvent idéale pour éviter de trop altérer le modèle d'origine.
  Un learning rate trop élevé risque de déformer le modèle de base, tandis qu'un taux trop bas rendrait l'apprentissage trop lent. Cette plage est un bon compromis.

**LR Scheduler – Constant with a Warmup of 0%**

- Un LR scheduler constant signifie que le learning rate reste fixe tout au long de l'entraînement. Pas de warmup signifie que le learning rate commence directement à la valeur spécifiée.
  Pour l'entraînement d'un style, un learning rate constant garantit un apprentissage stable et uniforme sur toute la durée des epochs.

**Optimizer – Adafactor**

- Adafactor est un optimiseur qui ajuste automatiquement la vitesse d'apprentissage en fonction des gradients, ce qui est utile pour les grands modèles.
  Cela aide à mieux gérer la mémoire tout en étant plus efficace pour entraîner des styles complexes sur des modèles pré-entraînés.

**Optimizer Extra Arguments**

- Ces arguments désactivent certains comportements par défaut d'Adafactor (comme le scaling automatique et le warmup).
  Ces ajustements rendent l'entraînement plus stable et contrôlé pour les petits datasets ou les cas spécifiques comme l'entraînement de styles.

**Max Resolution – 1024x1024 (ou 512 x 512 pour économiser la VRAM)**

- La résolution maximale correspond à la taille des images sur lesquelles le modèle est entraîné. Des résolutions plus élevées comme 1024x1024 permettent de capturer plus de détails dans les styles, dans notre cas nous avons utilisé des images de 512 x 512.
  En augmentant la résolution, tu captures plus de détails stylistiques, mais cela demande plus de mémoire GPU.

#### 4.4.1.2 Stable Diffusion Automatic1111 :

Stable Diffusion Automatic1111 est une interface utilisateur graphique (UI) populaire qui permet de gérer, tester et visualiser les modèles de génération d'images de manière interactive. Il s'agit d'un fork avancé de Stable Diffusion Web UI, offrant de nombreuses fonctionnalités pour expérimenter avec des modèles de diffusion de texte-à-image, comme Stable Diffusion. L'interface fournit des outils pour configurer facilement les paramètres de génération d'images, charger des modèles pré-entraînés ou des modèles LoRA, et ajuster des hyperparamètres.

Cette interface est principalement utilisée dans le cadre du projet ANIMO pour :

- Tester les modèles LoRA qui ont été entraînés avec Kohya_ss.
- Visualiser et ajuster les résultats en fonction des émotions détectées.
- Optimiser la génération d'images en temps réel avec des paramètres ajustables et des prévisualisations instantanées.

Une fois les modèles LoRA entraînés avec Kohya_ss, ils peuvent être chargés dans Stable Diffusion Automatic1111 pour générer des images adaptées aux émotions. L'interface nous permet de sélectionner le modèle LoRA et de l'appliquer lors de la génération des images.

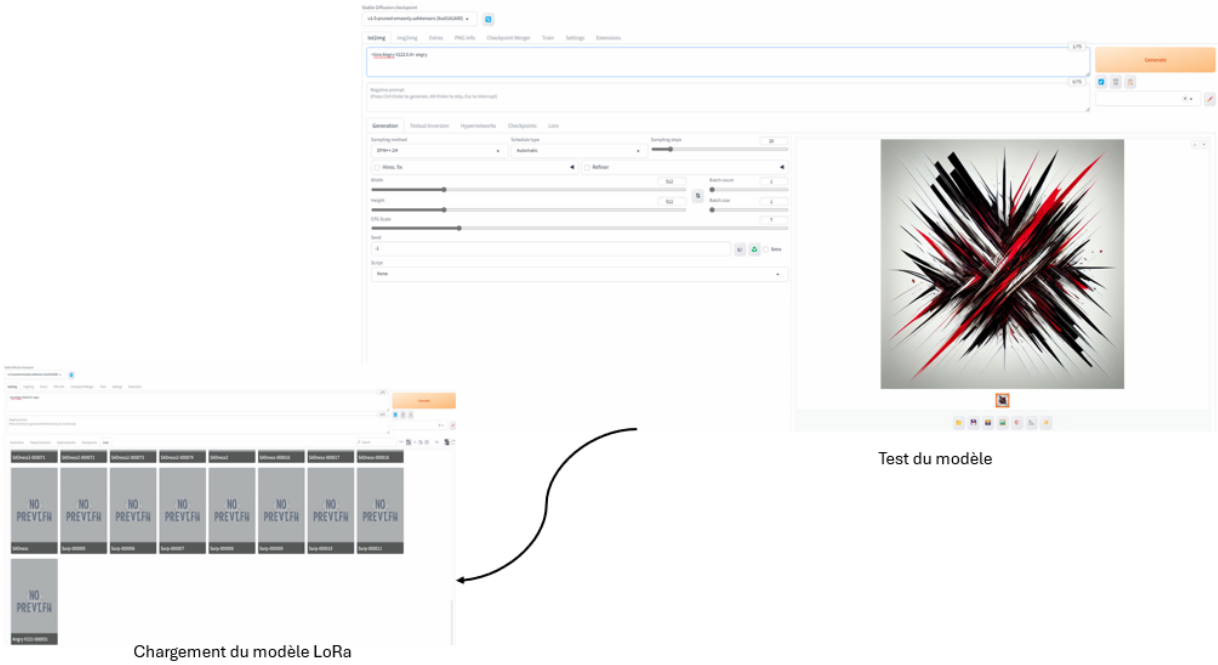



*Figure 37 : les étapes de test du modèle LoRa de l'émotion "Colère"*

#### 4.4.1.3 Evaluation des résultats :

L'évaluation des images générées dans le cadre du projet ANIMO repose sur un document fourni par l'artiste Lina KHEI, qui décrit les critères spécifiques pour chaque émotion. Ce document sert

de référence pour vérifier si les images générées respectent les caractéristiques visuelles associées à chaque émotion. Les critères d'évaluation incluent notamment les couleurs, le style des traits, et d'autres éléments visuels.

Pour chaque émotion détectée, le document de l'artiste spécifie des éléments clés que les images générées doivent respecter :

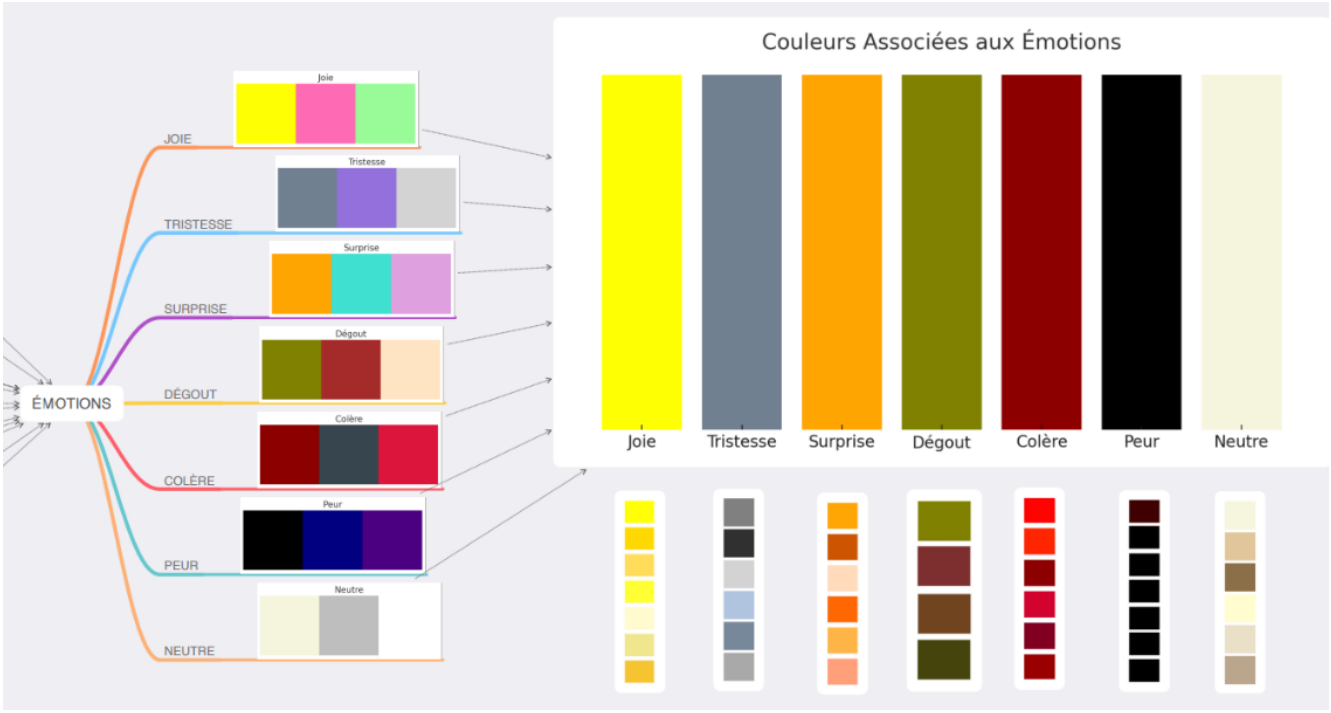



*Figure 38 : Carte heuristique de palette colorimétrique de chaque émotion*

Chaque émotion choisie a été associée à une palette colorimétrique distincte et à des éléments graphiques spécifiques afin de créer un langage visuel unique et expressif. Ces choix visent à évoquer de manière précise et nuancée, la profondeur des états émotionnels chez le public. Précisons également que nous pourrons y intégrer des images ou photos en cohérence avec les données rapportées.

**Joie :**

- **Palette colorimétrique :** Une palette de couleurs dominée par des teintes jaunes vives et éclatantes, évoquant la chaleur, l'énergie et la positivité. Des nuances d'orange et de vert lumineux peuvent également être intégrées pour ajouter de la vivacité.
- **Graphisme :** Les éléments graphiques comprennent des formes arrondies, des motifs dynamiques et des lignes courbes, qui évoquent l'énergie positive et la vitalité.

**Tristesse :**

- **Palette colorimétrique :** Une palette composée de teintes froide et neutre, rappelant la sérénité et la mélancolie. Des nuances de gris et de mauve peuvent être utilisées pour créer une ambiance délicate et apaisante.
- **Graphisme :** Les éléments graphiques sont délicats, avec des formes douces, des motifs fluides et des lignes douces, pour transmettre la douceur et la tranquillité de cette émotion.

**Peur :**

- **Palette colorimétrique :** Une palette sombre et intense comprenant des teintes noires, grises et des nuances profondes de pourpre ou de bleu foncé. Ces couleurs évoquent l'angoisse et la tension émotionnelle.
- **Graphisme :** Les éléments graphiques sont anguleux, avec des formes pointues, des motifs chaotiques et des lignes saccadées, pour créer une atmosphère angoissante et intense.

**Colère :**

- **Palette colorimétrique :** Une palette dominée par des teintes rouges vives et passionnées, représentant la force, la puissance et l'expression de la colère. Des touches de noir peuvent être utilisées pour renforcer l'intensité.
- **Graphisme :** Les éléments graphiques sont audacieux, avec des formes géométriques fortes, des motifs agressifs et des lignes percutantes, pour refléter la puissance et l'expression de la colère.

**Surprise :**

- **Palette colorimétrique :** Une palette contrastée et dynamique, mêlant des teintes vives comme le violet, l'orange et le turquoise. Ces couleurs captent l'effet de surprise et créent une atmosphère d'émerveillement.
- **Graphisme :** Les éléments graphiques sont soudains, avec des formes explosives, des motifs surprenants et des lignes expressives, pour capturer l'effet de surprise de cette émotion.

**Dégoût :**

- **Palette colorimétrique :** Une palette sombre et terne, comprenant des teintes vertes foncées, des nuances de brun et des touches de gris. Ces couleurs évoquent le dégoût et l'aversion, créant une atmosphère répugnante.
- **Graphisme :** Les éléments graphiques sont désordonnés, avec des formes irrégulières, des motifs répugnants et des lignes tortueuses, pour évoquer le dégoût et l'aversion.

www.linakheiart.com

*Figure 39 : Les critères des visuelles générées à respecter pour les émotions de base*

## *Exemples des visuelles générées avec les modèles LoRa entrainés :*

Extrait des images de la base de données

Les images générées

Peur

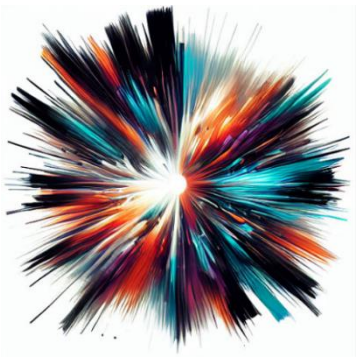
Dégout

Extrait des images de la base de données

Les images générées

Joie

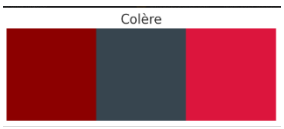
Surprise



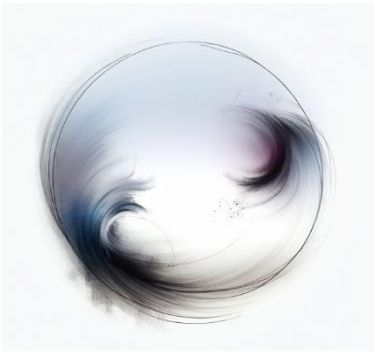
Extrait des images de la base de données

Les images générées

Colère



Tristesse



Extrait des images de la base de données  Les images générées

Neutre

## 4.5 Le flux de travail total de projet ANIMO :

### 4.5.1 Le pipeline de détection et Analyse d'émotions :

Le pipeline de détection et d'analyse d'émotions représente une avancée significative dans les applications interactives basées sur la reconnaissance faciale, en assurant non seulement une précision et une réactivité optimales, mais aussi une sécurité renforcée. Cette infrastructure technologique utilise des bibliothèques avancées telles que Dlib pour la détection de visage et les landmarks, ainsi que DeepFace pour l'analyse d'émotion, tout en intégrant des mesures de sécurité strictes pour protéger la confidentialité des utilisateurs.

Au cœur de ce système se trouve la capacité robuste de Dlib à détecter efficacement les visages dans des conditions variées, tout en préservant la sécurité des informations personnelles. Seuls les points clés du visage, tels que les landmarks représentant les yeux, le nez et la bouche, sont extraits et utilisés pour suivre et analyser les expressions faciales. Cette approche garantit que seules les données essentielles et non identifiables sont traitées, préservant ainsi la confidentialité des utilisateurs tout en permettant une interaction fluide et intuitive avec l'application.

L'analyse d'émotion, facilitée par DeepFace, enrichit cette expérience en identifiant et en catégorisant les émotions à partir des expressions faciales détectées, tout en respectant les principes de

confidentialité et de sécurité des données. Cette technologie spécialisée assure une réponse précise et rapide aux signaux émotionnels des utilisateurs, tout en limitant la collecte et l'utilisation des informations personnelles aux seuls besoins fonctionnels de l'application.

Pour optimiser les performances et garantir une réactivité maximale, le traitement parallèle avec des threads est employé, tout en intégrant des mécanismes de sécurité robustes. Ces mesures incluent l'utilisation de techniques de verrouillage pour sécuriser l'accès aux données sensibles partagées entre les threads, assurant ainsi une gestion sécurisée et cohérente des informations personnelles tout au long de l'interaction utilisateur.

Enfin, l'affichage des résultats en temps réel sur une interface graphique est soigneusement conçu pour ne montrer que les informations non identifiables et pertinentes pour l'expérience utilisateur, renforçant ainsi la confidentialité tout en offrant une interaction immédiate et immersive avec les données détectées.

En combinant ces composants technologiques avancés avec des pratiques de sécurité intégrées, le pipeline de détection et d'analyse d'émotions établit une norme pour les applications interactives basées sur la reconnaissance faciale, offrant à la fois précision, réactivité et confidentialité renforcée pour les utilisateurs.

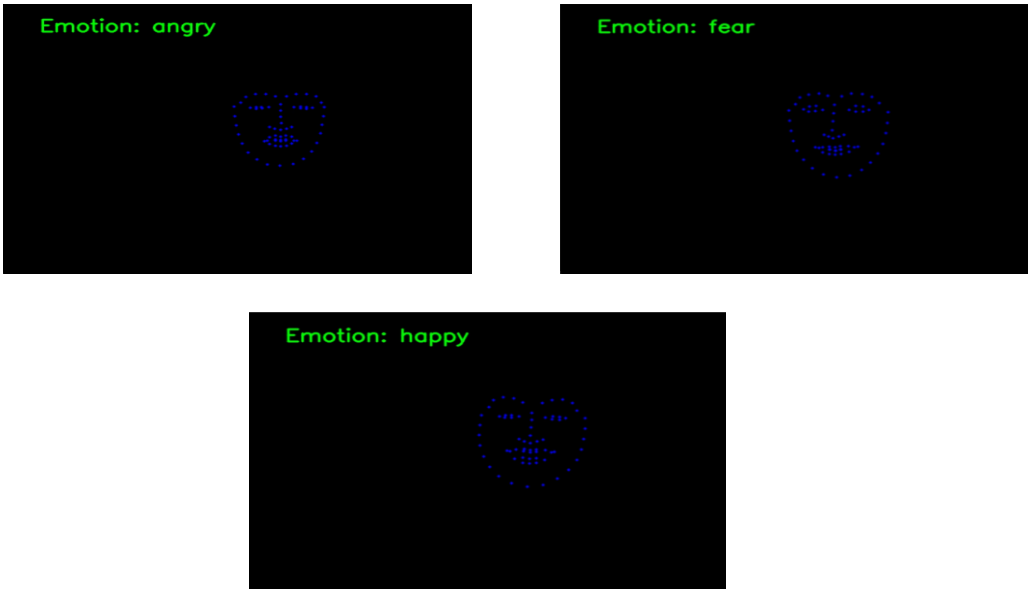



*Figure 40 : Système de détection des émotions*

### 4.5.2 Pipeline de Génération d'Images :

**Traitement de l'Émotion :** Utilisation de l'émotion détectée comme stimulus pour la génération d'images artistiques. Chaque émotion est associée à un modèle LoRA spécifiquement entraîné pour produire des œuvres visuelles correspondant à cette émotion particulière.

**Sélection du Modèle LoRA :** Sélection dynamique du modèle LoRA adapté à l'émotion détectée. Chaque modèle LoRA est préalablement configuré pour exprimer de manière optimale les nuances et les caractéristiques visuelles associées à chaque état émotionnel.

**Chargement et Génération d'Image :** Chargement des paramètres pré-entraînés du modèle LoRA sélectionné, suivi de l'utilisation de la librairie Stable Diffusion Pipeline pour générer une image basée sur le

prompt spécifique à l'émotion détectée. Cette étape intègre des techniques avancées de filtrage et d'amélioration de la qualité visuelle.

**Sauvegarde et Affichage de l'Image :** Enregistrement local de l'image générée et mise à jour de l'interface utilisateur pour afficher l'œuvre finale.

**Répétition du Processus :** Répétition continue du processus de génération d'images à chaque nouvelle détection d'émotion, assurant une expérience interactive et enrichissante pour les utilisateurs tout au long de leur session thérapeutique.

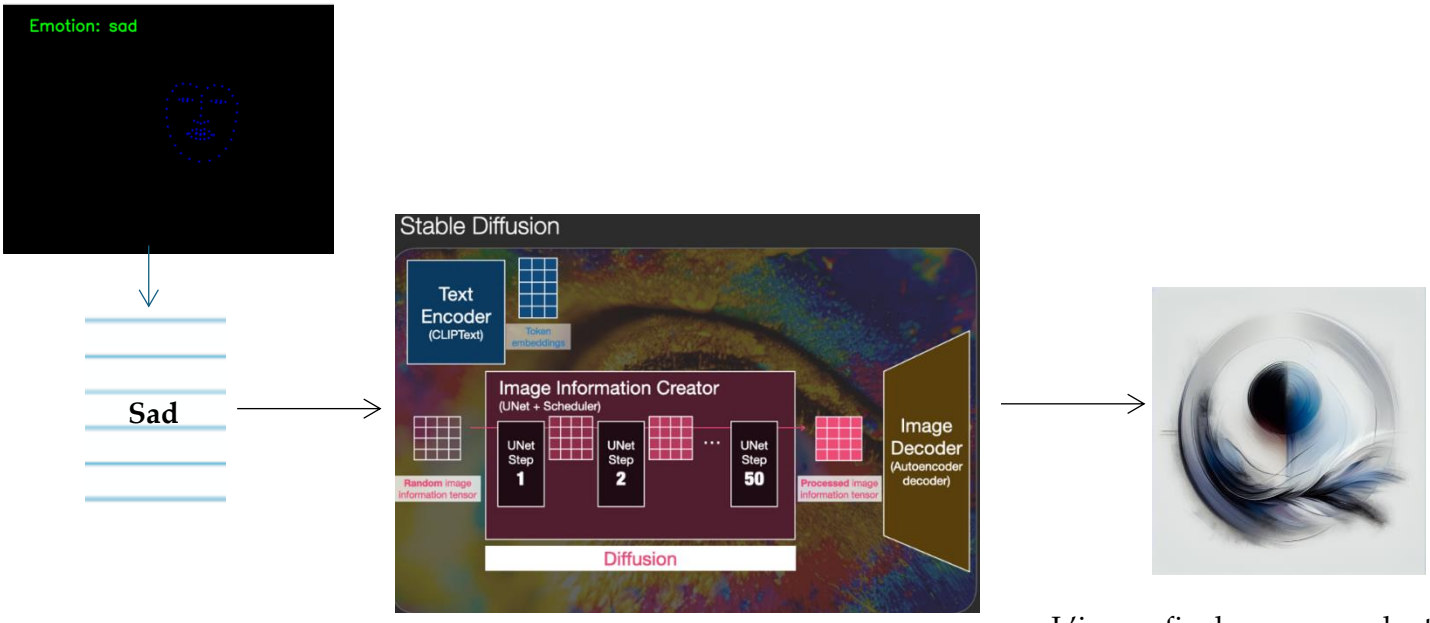



*Figure 41 : Le pipeline de génération d'images artistiques*

### 4.5.3 Interface graphique

#### 4.5.3.1 Technologies utilisées :

L'interface graphique du projet ANIMO est conçue pour permettre la génération d'images à partir de prompts textuels et la détection d'émotions. Elle est construite avec les technologies suivantes :

**FastAPI** : Un framework léger pour la création d'API web en Python. FastAPI est utilisé pour gérer la communication entre le backend et l'interface utilisateur, en permettant la soumission des prompts, le traitement des émotions détectées, ainsi que la gestion des fichiers et des images générées.

**WebSockets** : Utilisés pour la communication en temps réel entre le serveur et l'interface utilisateur, WebSockets permettent de mettre à jour instantanément les images générées sur l'interface en fonction des changements d'émotions détectées chez l'utilisateur.

**Socket.IO :** Cette bibliothèque facilite la gestion des WebSockets, permettant une communication bidirectionnelle en temps réel. Elle est utilisée pour envoyer les informations sur les nouvelles

images générées ou les positions de celles-ci sur la toile à l'interface utilisateur.

**Jinja2 :** Un moteur de templates HTML utilisé pour afficher dynamiquement les images et leurs positions sur la toile. Il permet également de gérer les sessions utilisateurs et de maintenir un historique des images générées.

**Bootstrap :** Le framework CSS est utilisé pour la mise en forme et la structuration de l'interface, garantissant une apparence moderne et réactive.

#### 4.5.3.2 Processus de l'interface graphique:

**Soumission de Prompt :** L'utilisateur entre un prompt dans le champ prévu à cet effet et soumet le formulaire. Le backend reçoit ce prompt, sélectionne un modèle LoRA spécifique en fonction de l'émotion ou du texte détecté dans le prompt, et utilise le modèle Stable Diffusion pour générer une image correspondante.

**Génération d'Image :** Le pipeline de Stable Diffusion est utilisé pour créer une image à partir du prompt soumis. Si une émotion est détectée ou mentionnée dans le prompt, un modèle LoRA correspondant est chargé pour ajuster la génération d'image selon l'émotion spécifiée.

**Affichage sur la Toile :** L'image générée est ensuite placée à une position spécifique sur la toile virtuelle, une zone de 720 x 985 pixels où plusieurs images peuvent être affichées ensemble. L'image est redimensionnée et insérée à une position calculée de manière aléatoire.

**Détection d'Émotions en Temps Réel :** L'utilisateur peut activer la détection des émotions en temps réel. Le modèle de détection d'émotions, intégré dans l'application, analyse les expressions faciales de l'utilisateur pour identifier une émotion dominante. L'émotion détectée est ensuite utilisée pour générer automatiquement une nouvelle image, qui est affichée sur la toile.

**Téléchargement des Œuvres :** Une fois que l'utilisateur est satisfait des images générées, il peut télécharger la toile combinée sous forme d'image unique en cliquant sur le bouton "Download Artwork".

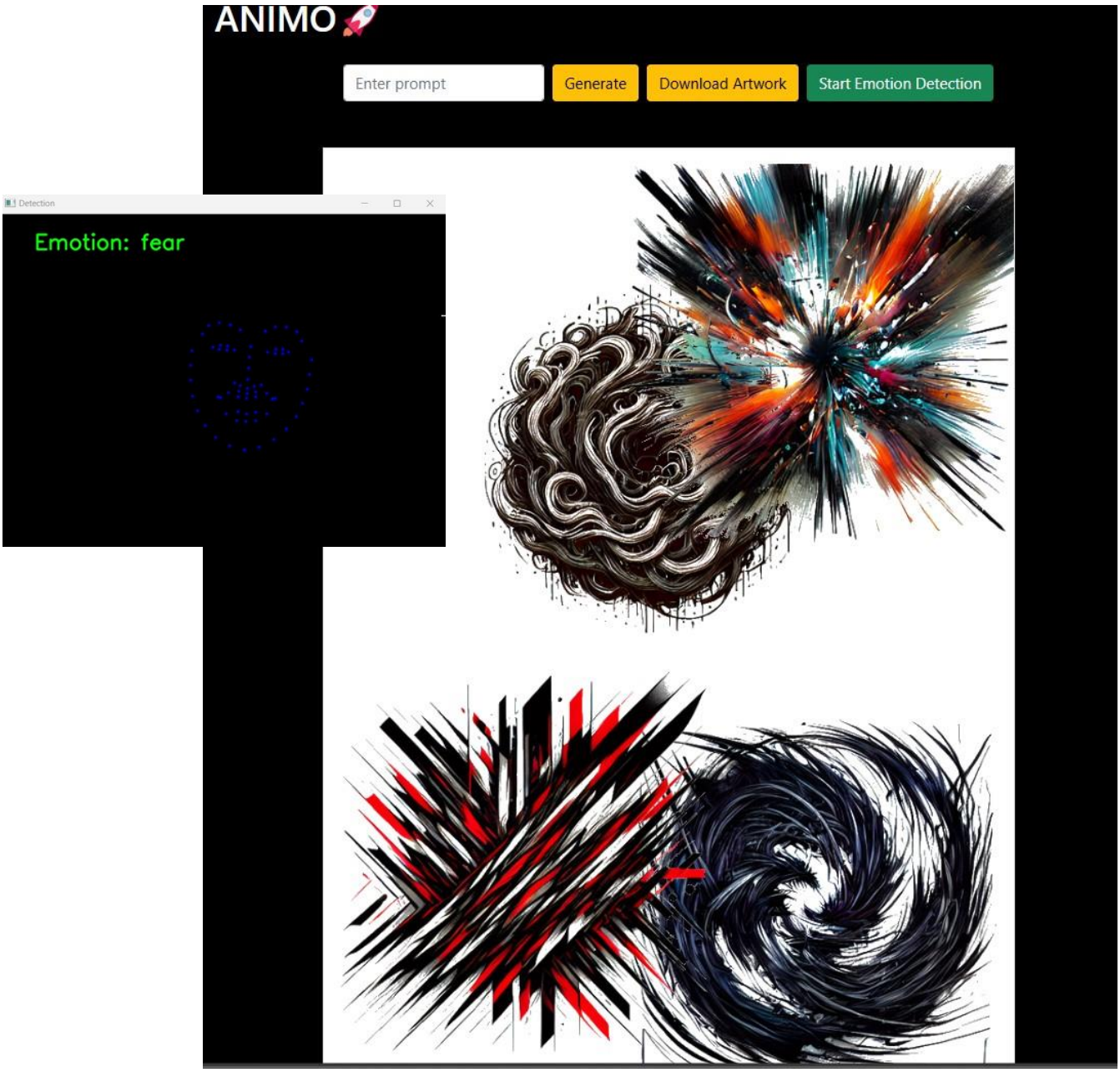



*Figure 42 : L'interface graphique ANIMO*

## 4.6 Conclusion :

L'intégration de Kohya_ss et de Stable Diffusion Automatic1111 constitue un flux de travail efficace pour l'entraînement, le test et la validation des modèles LoRA dans le projet ANIMO. Kohya_ss permet d'entraîner des modèles spécifiques aux émotions, tandis que Stable Diffusion Automatic1111 offre un environnement convivial pour tester et visualiser les résultats. Ce processus permet de générer des images en temps réel qui correspondent aux émotions détectées, garantissant une expérience utilisateur immersive et personnalisée.

# Conclusion Générale

Le projet ANIMO constitue une avancée majeure dans la fusion de l'art et de la thérapie, soutenu par les dernières innovations en intelligence artificielle et en reconnaissance faciale. Son objectif est de révolutionner l'art-thérapie en offrant une plateforme interactive et dynamique qui permet aux utilisateurs d'exprimer et de comprendre leurs émotions à travers la création artistique numérique en temps réel.

ANIMO se distingue par plusieurs innovations technologiques et conceptuelles. Tout d'abord, la détection des émotions en temps réel, réalisée grâce à des technologies comme Dlib pour la reconnaissance faciale et DeepFace pour l'analyse émotionnelle, permet une reconnaissance rapide et précise des expressions faciales. Cela garantit une expérience immersive où les émotions sont captées et interprétées instantanément. Ensuite, la génération d'images artistiques de haute qualité repose sur l'utilisation de Stable Diffusion et LoRA, permettant de créer des visuels qui s'adaptent dynamiquement aux émotions détectées. Cette interaction continue entre l'analyse émotionnelle et la création artistique renforce l'aspect personnalisé et interactif du projet.

La sécurité et la confidentialité des données des utilisateurs sont également au cœur d'ANIMO. Le projet veille à ce que seules les informations essentielles et non identifiables soient traitées, assurant ainsi une gestion sécurisée des données personnelles. En outre, grâce à l'intégration de modèles LoRA spécifiques à chaque émotion, ANIMO offre une personnalisation unique et une réactivité maximale, adaptant les visuels générés aux états émotionnels des utilisateurs pour une expérience fluide et sur mesure.

En dépassant les méthodes traditionnelles de l'art-thérapie, ANIMO propose une interaction directe avec l'utilisateur, en captant les émotions et en les transcrivant sous forme d'images dynamiques, évolutives en temps réel. Cette approche permet de fournir un support thérapeutique immédiat et personnalisé, rendant chaque session unique et adaptée aux besoins spécifiques des utilisateurs.

Bien que le projet ait atteint ses objectifs, plusieurs pistes d'amélioration sont envisagées. L'optimisation des temps de génération pourrait encore améliorer la fluidité de l'expérience, en réduisant le délai entre la détection des émotions et la création des visuels. L'intégration de nouvelles formes d'expression artistique, telles que la musique ou les animations 3D, pourrait également enrichir l'expérience et offrir une plus grande diversité émotionnelle. Enfin, ANIMO présente un fort potentiel d'adaptation à d'autres contextes thérapeutiques, tels que la gestion du stress ou l'accompagnement de patients souffrant de troubles anxieux ou dépressifs.

Le projet ANIMO représente une avancée majeure dans l'application de l'intelligence artificielle à des fins thérapeutiques. En alliant détection émotionnelle en temps réel et génération d'images dynamiques, il redéfinit la manière dont les émotions sont interprétées et traitées, tout en ouvrant des perspectives prometteuses pour l'avenir de l'art-thérapie et d'autres domaines de la santé mentale.

# Bibliographie:

Article: "Les Generative Adversarial Networks". 2022. ffhal-04140057 Moez Krichen. https://hal.science/hal-04140057/document

Article: "Perspectives for Generative AI-Assisted Art Therapy" for Melanoma Patients,Lennart Jütte,Ning Wang , Martin Steven and Bernhard Roth. https://www.researchgate.net/publication/383835409_Perspectives_for_Generative_AI-Assisted_Art_Therapy_for_Melanoma_Patients

"Creativity and Style in GAN and AI Art: Some Art-historical Reflections" Jim Berryman https://www.researchgate.net/publication/380265241_Creativity_and_Style_in_GAN_and_AI_Art_Some_Art-historical_Reflections

"Image Generation using Generative Adversarial Network and Stable Diffusion" Kunal Wagh https://www.researchgate.net/publication/379949785_Image_Generation_using_Generative_Adversarial_Network_and_Stable_Diffusion

"Computational Creativity: AI-Generated Art and Creative Applications" Sanjana Sunil Wankhede,Ashish Deharkar,Lowlesh Nandkishor Yadav https://www.researchgate.net/publication/382530505_Computational_Creativity_AI-Generated_Art_and_Creative_Applications

"Using LoRA in Stable Diffusion" par *Jason Brownlee* https://machinelearningmastery.com/using-lora-in-stable-diffusion/

" Stable Diffusion" par **Robin Rombach** https://github.com/CompVis/stable-diffusion?tab=readme-ov-file#stable-diffusion-v1

"What are Diffusion Models?" https://www.geeksforgeeks.org/what-are-diffusion-models/

« High-resolution image synthesis with Latent Diffusion Models, » par R. Rombach, A. Blattmann, D. Lorenz, P. Esser, et B. Ommer, arXiv.org, 13-Apr-2022. [En ligne]. Disponible: https://arxiv.org/abs/2112.10752.

« The Illustrated Stable Diffusion », The Illustrated Stable Diffusion - Jay Alammar - Visualizing machine learning one concept at a time. [En ligne]. Disponible : https://jalammar.github.io/illustrated-stable-diffusion/.

" Stable diffusion : High-resolution image synthesis with latent diffusion models | ML coding series," par A. Gordić, YouTube, 01-Sep-2022. [En ligne].
Disponible : https://www.youtube.com/watch?v=f6PtJKdey8E

" Intrinsic Dimensionality Explains the Effectiveness of Language Model Fine-Tuning" par Armen Aghajanyan, Luke Zettlemoyer, Sonal Gupta
Disponible : https://arxiv.org/abs/2012.13255

" Kohya's GUI" par bmaltais (https://github.com/bmaltais) https://github.com/bmaltais/kohya_ss

" Introduction to Diffusion Models for Machine Learning" https://www.assemblyai.com/blog/diffusion-models-for-machine-learning-introduction/

Le projet ANIMO, présenté dans ce rapport, est une application innovante qui génère des images artistiques en temps réel en fonction des émotions détectées. Le terme "animo", qui signifie "souffle d'âme" en latin, symbolise la profondeur émotionnelle de ce projet. En collaboration avec l'artiste française Lina Khei, le projet associe des techniques d'intelligence artificielle telles que Stable Diffusion et LoRA pour créer des visuels artistiques correspondant aux émotions de l'utilisateur. Le rapport détaille toutes les étapes de mise en œuvre, incluant la détection des émotions et l'entraînement de modèles spécifiques à chaque émotion (joie, tristesse, colère, surprise,neutre,dégout,peur).

Le système génère des images statiques, mais des perspectives d'amélioration incluent la création de visuels dynamiques qui évoluent selon l'état émotionnel de l'utilisateur, enrichissant ainsi l'expérience thérapeutique. La personnalisation des images en fonction des émotions pourrait également être adaptée à des environnements immersifs en réalité virtuelle, ouvrant de nouvelles voies d'application pour l'art-thérapie.

The ANIMO project, introduced in this report, is an innovative application that generates artistic images in real-time based on emotion detection. The word "animo," meaning "breath of soul" in Latin, reflects the emotional depth of this project. Collaborating with French artist Lina Khei, the project combines AI techniques like Stable Diffusion and LoRA to create visual artworks corresponding to users' emotions. The report details the implementation stages, including emotion detection and the training of emotion-specific models (joy, sadness, anger, surprise, etc.).

Currently, the system generates static images, but future improvements could include the development of dynamic visuals that evolve with the user's emotional state, enriching the therapeutic experience. There are also prospects for adapting these personalized images to immersive virtual reality environments, opening new applications for art therapy.