

CAHIER DES CHARGES

CREATIVE AI

RÉALISÉ PAR :

Abdelghani Bougalla
Boukhsim Nabil
ER-Rami Youssef
Mohamed Bendadi

LICENCE

OPTION: INGÉNIERIE
DES SYSTÈMES
INFORMATIQUES

1. CONTEXTE GÉNÉRAL

Le projet Creative AI s'inscrit dans le domaine de l'intelligence artificielle appliquée à la création de contenus multimédias. Avec l'essor des industries créatives (jeux vidéo, cinéma, marketing, éducation, accessibilité), la demande pour des outils automatisés de génération de contenus audio, visuels et vidéo croît rapidement. Les technologies d'IA, telles que la synthèse vocale, le clonage vocal, la génération d'effets sonores, et la création d'images/vidéos, permettent de produire des contenus personnalisés à faible coût et avec une grande flexibilité.

2. PROBLÉMATIQUE



La création de contenus multimédias (voix, effets sonores, images, vidéos) est souvent coûteuse, chronophage et nécessite des compétences techniques ou des ressources humaines importantes. Les solutions existantes présentent plusieurs limites :

- **Synthèse vocale** : Les voix manquent parfois de naturalité, et le support pour les langues moins courantes est limité.
- **Clonage de voix** : Les outils sont souvent complexes, soulèvent des problèmes éthiques (ex. : utilisation non autorisée de voix), ou nécessitent de grandes quantités de données audio.
- **Génération d'effets sonores** : Les solutions pour créer des effets sonores à partir de texte sont rares, souvent peu précises ou difficiles à intégrer dans des projets multimédias.

- **Génération d'images/vidéos** : Les outils actuels peuvent manquer de personnalisation, produire des résultats incohérents ou être difficiles à utiliser pour des non-techniciens.
- **Accessibilité** : Les plateformes combinant audio (voix, effets), images et vidéos dans une interface intuitive et unifiée sont rares, limitant leur adoption.

Creative AI vise à répondre à ces problèmes en proposant une plateforme unifiée, accessible, multilingue et personnalisable, intégrant la synthèse vocale, le clonage de voix, la génération d'effets sonores, d'images et de vidéos, avec une forte attention à la sécurité, l'éthique et l'évolutivité.

3. OBJECTIFS

Les objectifs de **Creative AI** sont précis et mesurables :

- **Automatisation de la création de contenus** : Permettre la génération automatique de voix, effets sonores, images et vidéos à partir de texte ou d'échantillons, réduisant le temps de production de 50 % par rapport aux méthodes manuelles.
- **Support multilingue** : Offrir une synthèse vocale et un clonage de voix dans au moins 30 langues, avec une précision phonétique de 95 % ou plus.
- **Qualité des contenus** : Produire des voix naturelles (score MOS – Mean Opinion Score – supérieur à 4/5), des effets sonores précis, des images nettes (résolution jusqu'à 4K) et des vidéos fluides (jusqu'à 60 FPS en 1080p ou 4K).
- **Accessibilité** : Développer une interface intuitive utilisable par des non-techniciens, avec un temps d'apprentissage inférieur à 10 minutes.
- **Conformité éthique et légale** : Assurer la conformité RGPD et mettre en place des mécanismes de consentement pour le clonage de voix, avec 100 % des données sensibles chiffrées.

4. MOTIVATION

Creative AI est pertinent pour plusieurs raisons :

- **Réduction des coûts** : Automatiser la création de contenus multimédias réduit les dépenses liées à l'embauche de voix off, de sound designers, d'artistes ou de vidéastes.
- **Accessibilité accrue** : Les personnes malvoyantes, les créateurs à budget limité ou les petites entreprises peuvent utiliser des outils abordables pour produire des contenus de qualité.
- **Flexibilité** : Les entreprises (marketing, jeux vidéo) et les éducateurs peuvent personnaliser rapidement des contenus pour des publics variés.
- **Innovation** : En combinant synthèse vocale, clonage de voix, génération d'effets sonores, images et vidéos dans une plateforme unique, Creative AI se distingue des solutions fragmentées existantes.

5. BESOINS FONCTIONNELS ET NON FONCTIONNELS



5.1. BESOINS FONCTIONNELS

- **Synthèse Vocale (TTS) :**
 - Convertir du texte en voix naturelle en utilisant le modèle StyleTTS2.
 - Exporter les fichiers audio en MP3/WAV.
- **Clonage de Voix :**
 - Créer des voix synthétiques à partir d'échantillons audio en utilisant le modèle Seed-VC.
- **Génération d'Effets Sonores :**
 - Générer des effets sonores à partir de descriptions textuelles (ex. : "vent qui souffle", "explosion") en utilisant le modèle Make-An-Audio.
 - Exporter en MP3/WAV.
- **Génération d'Images :**
 - Produire des images à partir de prompts textuels avec des styles variés (réaliste, cartoon, abstrait) et résolutions jusqu'à 4K.
 - Exporter en PNG/JPEG.
- **Génération de Vidéos :**
 - Créer des vidéos à partir de texte, images ou audio (voix/effets sonores), avec narration vocale et animations simples.
 - Supporter 1080p/4K, exporter en MP4/MOV.
- **Authentification des Utilisateurs :**
 - Permettre l'inscription/connexion via email ou OAuth (Google, GitHub) avec NextAuth.js.
- **Gestion des Ressources :**
 - Tableau de bord pour gérer les voix, effets sonores, images et vidéos.

5.2. BESOINS NON FONCTIONNELS



- **Performance :**
 - Qualité audio (MOS > 4/5), effets sonores précis, images nettes, vidéos fluides.
- **Sécurité :**
 - Chiffrement AES-256 pour les données au repos, HTTPS pour les communications.
 - Conformité RGPD, avec consentement explicite pour le clonage de voix.
 - Authentification sécurisée via NextAuth.js.
- **Accessibilité :**
 - Interface conforme WCAG 2.1, compatible avec les lecteurs d'écran.
- **Fiabilité :**
 - Disponibilité de 99,9 % (SLA), avec journalisation via Inngest pour diagnostiquer les erreurs.
- **Portabilité :**
 - Compatible avec les navigateurs modernes (Chrome, Firefox, Safari) et responsive pour mobile/tablette via Tailwind.

6. TECHNOLOGIES UTILISÉES

Composant	Technologie	Justification
Modèle TTS	StyleTTS2	Modèle avancé pour des voix naturelles, multilingue, personnalisation émotionnelle.
Clonage de Voix	Seed-VC	Efficace pour le clonage avec peu de données, support multilingue.
Effets Sonores	Make-An-Audio	Génère des effets précis à partir de texte, adapté aux besoins créatifs.
Génération d'Images	Stable Diffusion	Produit des images variées et de haute qualité, largement adopté.
Génération de Vidéos	Replicate, GeminiAi	Supporte la génération vidéo
Framework IA	PyTorch	Flexible, optimisé pour les réseaux neuronaux, supporte StyleTTS2/Seed-VC.
Backend	FastAPI (Python)	Performant, idéal pour les services backend, intégration avec PyTorch.
Tâches Asynchrones	Inngest	Gestion efficace des tâches lourdes (clonage, génération vidéo).
Base de Données	Prisma, Drizzle	Prisma pour une ORM robuste, Drizzle pour des performances légères et flexibles.
Frontend	Next.js 15, React, Tailwind	Next.js pour SSR/SSG, React pour interactivité, Tailwind pour un design rapide.

6. TECHNOLOGIES UTILISÉES

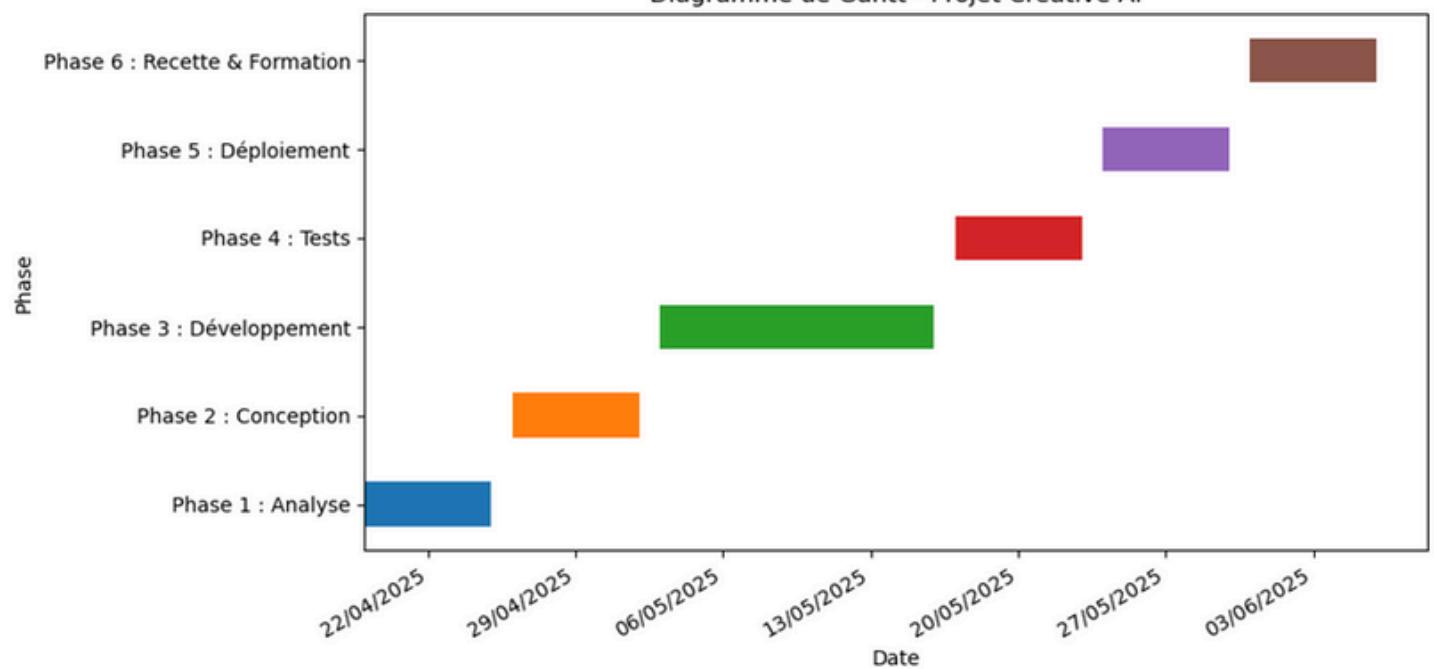
Composant	Technologie	Justification
Authentification	NextAuth.js	Simplifie l'authentification OAuth/email, sécurisé et intégré à Next.js.
Communication	tRPC	Type-safe, performant pour la communication frontend-backend.
Stockage	AWS S3	Évolutif, sécurisé, adapté au stockage de fichiers audio/images/vidéos.
Cloud	AWS (EC2, S3, ECS)	Évolutif, fiable, supporte GPU pour l'inférence IA.
Conteneurisation	Docker	Environnements cohérents, déploiement simplifié.



7. PLANIFICATION

Phase	Début	Fin	Jalons clés
Analyse	19/04/2025	25/04/2025	Rédaction du CDC, spécifications détaillées
Conception	26/04/2025	02/05/2025	Architecture, prototypes UI
Développement	03/05/2025	16/05/2025	Implémentation API, modules IA
Tests	17/05/2025	23/05/2025	Tests unitaires, intégration, performance
Déploiement	24/05/2025	30/05/2025	Remise du rapport final
Recette & Formation	31/05/2025	06/06/2025	Formation des utilisateurs

Diagramme de Gantt - Projet Creative AI



8. ÉTUDE BENCHMARKING

Cette section compare **Creative AI** à des solutions existantes dans les domaines de la synthèse vocale, du clonage de voix, de la génération d'effets sonores, d'images et de vidéos.

Solution	Forces	Faiblesses	Différenciation Creative AI
Solution TTS	Qualité vocale élevée, support multilingue.	Voix moins naturelles, personnalisation limitée pour langues rares.	StyleTTS2 pour voix ultra-réalistes.
Solution Clonage	Clonage précis avec données limitées.	Problèmes éthiques, interface complexe.	Seed-VC avec consentement explicite, interface intuitive.
Solution Effets Sonores	Effets variés, intégration audio.	Précision limitée, manque de personnalisation.	Make-An-Audio pour effets précis, ajustements faciles.
Solution Images	Images de haute qualité, styles variés.	Temps de génération long, options limitées.	Stable Diffusion, génération rapide, personnalisation.
Solution Vidéos	Animations fluides.	Résolutions limitées, narration vocale absente.	VideoGPT, support 4K