**Plan de Mitigation Technique**

**A. Données et Prétraitement**

Un pipeline de nettoyage automatique est mis en place pour garantir la qualité et la sécurité des données. Il comprend la suppression des métadonnées EXIF, le floutage des visages non consentis et l’élimination des données sensibles. Toutes les images et textes sont anonymisés avant tout usage dans les modèles d’apprentissage. Les environnements de travail sont strictement séparés : les jeux d’entraînement, les bases clients et les logs sont isolés afin d’éviter toute contamination croisée. Une documentation complète du dataset (DataSheet) décrit l’origine des données, les biais potentiels et les méthodes de sélection. Des scripts automatisés en Python filtrent les variables sensibles avant tout traitement.

**B. Développement du Modèle**

L’explicabilité (XAI) est intégrée à travers des outils comme LIME et SHAP pour interpréter les recommandations générées. L’application affiche la logique de suggestion à l’utilisateur, par exemple “basé sur vos styles précédents et vos préférences déclarées”. Une supervision humaine est obligatoire avant toute décision de refus automatisé, notamment sur l’âge estimé ou le scoring psychologique. Chaque itération du modèle fait l’objet d’une validation technique continue incluant des tests unitaires et de cohérence. Un audit interne pré-déploiement évalue les biais, la précision et les taux d’erreurs discriminantes.

**C. Infrastructure et Sécurité**

L’architecture est sécurisée de bout en bout. L’authentification s’effectue via OAuth2 pour toutes les API IA, et les communications sont chiffrées en TLS. Les modèles et datasets sont stockés sous chiffrement AES-256. Une journalisation complète assure la traçabilité : chaque requête IA (entrée, sortie, version du modèle, timestamp) est enregistrée pour audit ou révision. Un monitoring en temps réel surveille la dérive des prédictions, les anomalies et le taux d’erreurs pour garantir la fiabilité du système.

**D. Gouvernance Modèle et MLOps**

Une approche MLOps by design est adoptée pour chaque modèle avec un pipeline CI/CD structuré. Les étapes obligatoires incluent la validation des données, la vérification des biais, les tests fonctionnels et de sécurité, le versioning via MLflow ou DVC, et un déploiement contrôlé de l’environnement de staging à la production. Le contrôle de dérive (concept drift) est assuré par une surveillance continue des écarts entre les prédictions et la réalité observée. Des scripts génèrent automatiquement la documentation du modèle sous forme de DataCard ou ModelCard.

**E. Exploitation et Maintenance**

Les données d’entraînement sont purgées de manière planifiée, par exemple tous les douze mois, afin de respecter les principes de minimisation et d’actualisation. Un audit trimestriel IA vérifie les performances, les biais et la conformité du système. Une formation continue des data engineers garantit la maîtrise des exigences de l’IA Act en matière de transparence, de traçabilité et d’éthique.

**F. Architecture Technique Simplifiée**

Frontend sécurisé (consentement utilisateur)

↓

API Gateway chiffrée (OAuth2 + logs)

↓

Data Processing Pipeline (nettoyage, anonymisation)

↓

Model Training & Registry (MLflow / DVC)

↓

Explainable AI Module (LIME/SHAP)

↓

Monitoring & Audit Service (logs + drift detection)

↓

Chiffrement & Backup sécurisé