



Mesures pour l'identification du contenu généré par l'Intelligence Artificielle et norme nationale obligatoire correspondante

2025.03.26

1. Contexte général

Dans le cadre de la mise en place d'un écosystème de contrôle fiable de l'Intelligence Artificielle (IA), l'Administration chinoise du cyberspace (CAC), le ministère de l'Industrie et des Technologies de l'information (MIIT), le ministère de la Sécurité publique (MPS) et l'Administration nationale de la radio et de la télévision (NRTA) ont promulgué les « *Mesures d'identification des contenus générés par l'IA*¹ » (les « Mesures »). Ces Mesures, ainsi que la norme nationale obligatoire GB 45438-2025 intitulée « *Technologie de cybersécurité : méthode d'identification des contenus générés par l'IA*² » (la « Norme ») entreront en vigueur le 1^{er} septembre 2025.

Les Mesures et la Norme permettent d'établir un cadre d'identification complet couvrant l'ensemble du cycle de vie du contenu généré par l'IA, de la production à la diffusion et à l'utilisation. Elles marquent ainsi un changement dans le contrôle du contenu généré par l'IA en Chine, qui passe d'une simple conformité formelle à une fiabilité technique.

2. Points clés

2.1 Champ d'application

Les fournisseurs de services d'information en ligne qui relèvent du « *Règlement sur la gestion des recommandations algorithmiques*³ », du « *Règlement sur la gestion de la synthèse profonde*⁴ » et des « *Mesures provisoires pour les services d'IA générative*⁵ » sont directement concernés par ces Mesures.

Tout le contenu généré par l'IA, qu'il soit sous forme de texte, image, son, vidéo, environnement virtuel ou toute autre information synthétique, rentre dans le champ d'application des Mesures.

¹ 人工智能生成合成内容标识办法

² GB 45438-2025 《网络安全技术 人工智能生成合成内容标识方法》

³ 算法推荐管理规定

⁴ 深度合成管理规定

⁵ 生成式 AI 服务暂行办法

2.2 Système de double étiquetage

Type d'Étiquette	Fonction	Exigences Techniques	Exemples
Étiquette Explicite	Indique clairement l'origine du contenu	- Intégrée de manière visible dans le contenu ou sur l'interface - Supporte une présentation multimodale (texte, graphique, audio)	Filigrane vidéo, superposition de textes, avertissement audio
Étiquette Implicite	Assure la traçabilité et la lutte contre la contrefaçon	- Intégrée dans les métadonnées du fichier - Stockée à l'aide de technologies anti-falsification	Données EXIF, certification blockchain

2.3 Exigences de conformité tout au long du cycle de vie du contenu

(1) Étape de production

Ainsi que résumé dans le tableau ci-dessus, pour tout contenu généré par l'IA, les fournisseurs de services doivent ajouter :

- Une étiquette explicite, sous forme de texte, audio, image, vidéo, scène virtuelle, etc... précisant clairement que ce contenu a été généré par l'IA (par exemple : la mention « générée par l'IA » est insérée sur la première image d'une vidéo).
- Une étiquette implicite, permettant d'enregistrer les métadonnées, telles que l'horodatage de génération du contenu, la version du modèle utilisé, l'entité responsable, etc.

(2) Étape de diffusion

- Les plateformes de contenu doivent vérifier les étiquettes implicites et s'assurer de la présence de mentions explicites lors de la publication du contenu (par exemple : "ce contenu contient des éléments générés par l'IA").
- Les plateformes de distribution doivent, lors de l'approbation du contenu à diffuser, s'assurer que la technologie d'étiquetage utilisée par les fournisseurs de services est bien conforme aux exigences réglementaires.

(3) Étape d'utilisation

- Les utilisateurs qui chargent du contenu généré par l'IA doivent le déclarer proactivement et utiliser les outils d'étiquetage fournis par la plateforme.
- La falsification, la suppression ou le contournement des mesures d'identification est strictement interdit.

2.4 Points clés de la Norme (GB 45438-2025)

(1) La Norme impose un format spécifique pour la taille/position de la police de l'étiquette explicite, ainsi que pour le contenu des métadonnées de l'étiquette implicite (les champs

obligatoires incluant le créateur du contenu, l'ID du modèle, l'horodatage).

(2) Elle impose aussi l'utilisation d'un mécanisme anti-falsification destiné à assurer l'intégrité des étiquettes, tel que le hachage cryptographique ou la signature numérique.

(3) Elle impose enfin aux plateformes de fournir une API pour permettre aux agences de régulation de vérifier les étiquettes.

2.5 Calendrier de mise en conformité

Délais	Actions clés pour la mise en conformité
Avant le 30 juin 2025	Adapter la technologie d'étiquetage et mettre à jour les politiques internes
À partir du 1er septembre 2025	Mettre en œuvre pleinement le système de double étiquetage, et s'assurer de la pleine conformité à toutes les étapes.

3. Conclusion et recommandations pratiques

Les Mesures et la Norme introduisent des obligations techniques et réglementaires strictes pour les entreprises opérant dans l'écosystème des contenus générés par l'IA. Pour garantir une transition fluide avant la date limite de septembre 2025, nous conseillons aux entreprises de mettre en œuvre de manière proactive des mécanismes d'étiquetage robustes, de mettre à jour les cadres de conformité et de travailler avec des organismes technologiques qualifiés.

Plus précisément, les actions suivantes peuvent être accomplies :

3.1 Mise en œuvre technique

(1) Déployer des outils d'étiquetage automatisés

→ Intégrer des SDKs/APIs conformes à la Norme (tel que *AIGC Tagger* de Baidu ou *MetaWatermark* d'Alibaba) permettant d'automatiser l'intégration des étiquettes explicites/implicites.

→ Pour les étiquettes explicites, adopter des formats spécifiques aux plateformes :

- **Texte** : Insérer des en-têtes/pieds de page (par exemple : "Généré par l'IA").
- **Images** : Ajouter des filigranes d'angle (par exemple : un logo translucide "Généré par l'IA").
- **Vidéo** : Inclure un avertissement superposé de 3 secondes (par exemple : "Cette vidéo contient du contenu généré par l'IA").
- **Audio** : Intégrer des annonces vocales (par exemple, "Cet audio a été synthétisé à l'aide de l'IA").

→ Pour les étiquettes implicites, utiliser des modèles de métadonnées avec des champs obligatoires : créateur, ID du modèle, horodatage.

(2) Mettre en œuvre des solutions anti-falsification :

→ Utiliser une certification basée sur les blockchain (tel que *ChainTrust* de Tencent) ou des horodatages numériques conformes à la RFC 3161 pour l'intégrité des métadonnées.

3.2 Ajustement opérationnel

- Modifier les pipelines de création de contenu pour inclure l'étiquetage, assurant la conformité à chaque étape.
- Effectuer une revue manuelle secondaire pour le contenu à haut risque (comme les scènes virtuelles et les portraits synthétiques de figures politiques).

3.3 Mises à jour des accords et de la politique de confidentialité

- Introduire une clause d'identification du contenu généré par l'IA dans les accords utilisateurs, décrivant, notamment, les obligations des utilisateurs en matière de conformité des étiquettes.
- Clarifier la portée et l'objectif de la collecte de métadonnées dans les politiques de confidentialité pour atténuer les risques de réclamations de collecte excessive de données.