Modalités d'utilisation d'Albert API

1. Champ d'application

Le présent document définit les modalités d'utilisation d'Albert API, la plateforme d'inférence d'intelligence artificielle (IA) générative interministérielle conçue pour le système d'information et de communication de l'Etat au sens de l'article 1 du décret n° 2019-1088 du 25 octobre 2019.

2. Objet de la plateforme

Le service permet aux équipes de développement informatique des administrations relevant du système d'information et de communication de l'Etat d'avoir accès à une interface de programmation (API) permettant d'interagir avec des modèles d'IA, notamment générative. Cette API permet d'intégrer des fonctionnalités utilisant l'IA générative dans les produits numériques de ces équipes.

Albert API est développé et opéré par la Direction interministérielle du numérique (DINUM). Toute utilisation de ce service doit respecter les présentes modalités d'utilisation et, le cas échéant, la doctrine d'emploi de votre administration de rattachement.

Le service est mis à disposition par la DINUM aux administrations relevant du système d'information et de communication de l'Etat, à destination de leurs agents et prestataires invités par celles-ci.

3. Définitions

Utilisateur : toute personne physique travaillant pour une administration de rattachement, qu'elles soient des agents ou des prestataires de celles-ci.

Administration de rattachement : une personne morale relevant système d'information et de communication de l'Etat au sens de l'article 1er du décret n° 2019-1088 du 25 octobre 2019qui met à disposition Albert API à ses agents ou prestataires.

Interface de programmation (API) : interface logicielle qui permet à une équipe de développement informatique d'accéder aux fonctions d'une application et de développer de nouvelles applications utilisant ces fonctions (JORF n°0101 du 30 avril 2022).

Modèle de langage ou « LLM » : algorithme d'intelligence artificielle composé d'un réseau de neurones artificiels entrainé sur de vastes quantités de données. Ces modèles se caractérisent par leur capacité à générer à partir d'une séquence en langage à naturel (ex : question en français) à générer une nouvelle séquence originale cohérente avec celle fournie en entrée (ex : réponse à la question). Ces modèles et leurs performances évoluent constamment et sont devenus très populaires depuis 2022.

Inférence : l'étape où un modèle d'IA, après avoir été entraîné sur un ensemble de données, est appliqué à de nouvelles données pour produire des résultats ou des prédictions.

Intelligence artificielle : champ interdisciplinaire théorique et pratique qui a pour objet la compréhension de mécanismes de la cognition et de la réflexion, et leur imitation par un dispositif matériel et logiciel, à des fins d'assistance ou de substitution à des activités humaines (JORF n°0285 du 9 décembre 2018).

Intelligence artificielle générative : branche de l'intelligence artificielle mettant en œuvre des modèles génératifs, qui vise à produire des contenus textuels, graphiques ou audiovisuels (JORF n°0212 du 6 septembre 2024)..

Poids ouverts ou openweights : caractéristique d'un modèle à rendre public les valeurs des neurones artificiels qui le compose. Ces poids ainsi que l'architecture du modèle sont en libre disposition selon les modalités de licence que prévoit le fournisseur du modèle.

Fournisseur de modèles : entité privée ou publique ayant conçue et entrainée des modèles d'intelligence artificielle.

Token : dans le cadre d'un modèle de langage, un *token* est l'unité de texte utilisée par le modèle pour comprendre et générer du langage. Un *token* peut correspondre à un mot entier, une partie de mot ou un signe de ponctuation. Le modèle traite les textes en les découpant en *tokens* plutôt qu'en mots complets.

Génération augmentée par récupération (« Retrievial augmented generation » ou RAG) : technique d'optimisation de réponses de modèle d'intelligence artificielle générative. Cette méthode permet notamment d'améliorer la qualité de réponses aux requêtes en permettant aux grands modèles de langage (LLM) d'exploiter des ressources de données supplémentaires sans ré-entraînement.

Infrastructure « SecNumCloud » : infrastructure d'informatique en nuage (« cloud ») répondant au référentiel d'exigences relatif aux prestataires de services d'informatique en nuage (« SecNumCloud ») définit par l'ANSSI, garantissant un standard de sécurité de ladite infrastructure.

4. Accès au service

Albert API est gratuitement mis à disposition des utilisateurs ou utilisatrices par leur administration de rattachement.

Cette mise à disposition gratuite est à des fins d'expérimentation et est soumise à des limitations correspondant à l'usage visé. La DINUM se réserve l'appréciation de la définition de ces limites.

La DINUM se réserve également le droit de revoir à la hausse ces limitations après étude de l'usage fait. Cette étude peut être demandée auprès de l'équipe responsable d'Albert API à l'adresse albert.api@numerique.gouv.fr.

Sous certaines conditions, un accès privilégié au service Albert API peut être obtenu par convention entre l'administration demandeuse et la DINUM.

5. Fonctionnalités

5.1. Création de compte

Dans le cas où l'utilisateur ou utilisatrice ne créé pas son compte en utilisant ProConnect, il ou elle veille à choisir un mot de passe qui ne peut pas être deviné ou retrouvé par quelqu'un d'autre. Le mot de passe doit être propre au compte Albert API et ne pas être utilisé sur d'autres applications.

L'utilisateur ou utilisatrice s'inscrit sur Albert API en créant un profil. Cette inscription est personnelle. L'utilisateur ou l'utilisatrice, dans la section intitulée « Account », pourra générer des clefs d'accès API lui permettant d'appeler les modèles d'IA.

5.2. Accès à des modèles d'IA

Albert API met à disposition un accès à des modèles d'IA, notamment générative, dits « à poids ouverts » (« openweights ») hébergés sur une infrastructure SecNumCloud.

Albert API offre un accès à une gamme variée de modèles, correspondant à des cas d'usage et performances différents. La DINUM se réserve le droit de changer les modèles mis à dispositions.

La liste des modèles disponibles ainsi que les licences sous lesquels ils sont soumis est disponible à l'adresse suivante : https://albert.sites.beta.gouv.fr/solutions/models/

5.3. Gestion de comptes et des appels

Les utilisateurs et utilisatrices ont accès sur leur espace :

- au suivi des requêtes à l'API,
- au nombre total de requêtes,
- au nombre de *token* consommé,
- à la consommation carbone,

- à leur clé d’accès d’API et sa date d’expiration,
- le taux de limite pour chaque modèle de langage.

5.4. Ingestion de fichiers et moteur de recherche

Albert API met à disposition un accès à une base de données pour stocker des données textuelles ainsi qu'un moteur de recherche sur cette base de données. Ces fonctionnalités permettent aux utilisateurs de réaliser du RAG.

Les utilisateurs et utilisatrices peuvent intégrer de nouvelles données dans la base de données. Ces données ne sont pas partagées avec d'autres.

6. Engagements et responsabilités des utilisateurs et utilisatrices

6.1. Usages conformes

Le service est mis à disposition pour développer et faciliter l'intégration des fonctionnalités d'IA, notamment générative, au sein des produits informatiques de l'État.

L'utilisateur ou l'utilisatrice est responsable des données ou contenus qu'il ou elle utilise à travers Albert API. Il ou elle veille à n'intégrer que des données appropriées.

Les données envoyées via les fonctionnalités d'ingestion de fichier, préalables à la fonctionnalité de moteur de recherche, sont amenées à être stockées par le système Albert API.

L'utilisateur ou l'utilisatrice est responsable des données qu'il ou elle envoie pour stockage dans le système Albert API.

L'utilisateur ou l'utilisatrice s'assure de garder confidentielles ses clefs d'API. Toute divulgation de ses clefs, quelle que soit sa forme, est interdite. Il ou elle assume les risques liés à l'utilisation de son adresse courriel et de ses clefs d'API.

L'utilisateur ou l'utilisatrice est responsable de l'utilisation du contenu renvoyé par Albert API. En particulier, la technologie des modèles de langage ne saurait garantir une fiabilité parfaite de l'information. L'utilisateur ou l'utilisatrice s'engage à prendre les précautions nécessaires vis-à-vis des informations fournies en réponse aux requêtes effectuées à l'API.

6.2. Usages interdits

L'utilisateur ou l'utilisatrice s'engage à ne pas utiliser Albert API contrairement aux dispositions légales et réglementaires en vigueur.

En particulier, l'utilisateur ou l'utilisatrice s'assure, en amont de l'utilisation de l'API, que cette dernière respecte le cadre juridique en vigueur notamment le RGPD, la loi Informatique et libertés, le code de la propriété intellectuelle et le CRPA.

Albert API n'a pas vocation à recevoir des données identifiées comme sensibles, notamment au sens des articles 9 et 10 du règlement général sur la protection des données.

7. Engagements et responsabilités de la DINUM

7.1. Sécurité et accès à la Plateforme

La DINUM s'engage à la sécurisation du service, notamment en prenant toutes les mesures nécessaires permettant de garantir la sécurité et la confidentialité des informations fournies. La DINUM s'engage à fournir les moyens nécessaires et raisonnables pour assurer un accès continu au service.

La DINUM se réserve le droit de faire évoluer, de modifier ou de suspendre, sans préavis, le service pour des raisons de maintenance ou pour tout autre motif jugé nécessaire.

La DINUM se réserve le droit de suspendre ou supprimer un compte utilisateur ou utilisatrice du service qui aurait méconnu les présentes modalités d'utilisation, sans préjudice des éventuelles actions en responsabilité pénale et civile qui pourraient être engagées à l'encontre de l'utilisateur ou l'utilisatrice.

7.2. Open source et licences

Le code source de la Plateforme est libre et disponible ici : https://github.com/etalab-ia/OpenGateLLM[1].

Les contenus proposés par la DINUM sont sous Licence Ouverte, à l'exception des logos et des représentations iconographiques et photographiques pouvant être régis par leurs licences propres.

Les modèles et les fonctionnalités peuvent être soumises à leur propre licence. Dans ce cas de figure, la mention en est portée par la DINUM lors de leur utilisation. L'utilisateur ou l'utilisatrice est responsable du respect des conditions de cette licence[2].

8. Évolution des modalités d'utilisation

Les termes des présentes modalités d'utilisation peuvent être modifiés ou complétés à tout moment, sans préavis, en fonction des modifications apportées au service, de l'évolution de la législation ou pour tout autre motif jugé nécessaire. Ces modifications et mises à jour s'imposent à l'utilisateur ou l'utilisatrice qui doit, en conséquence, se référer régulièrement à cette rubrique pour vérifier les modalités en vigueur.

---

[1] Voir la licence du code d'Albert API : https://github.com/etalab-ia/OpenGateLLM/blob/main/LICENSE

[2] Voir les licences des modèles mis à disposition : https://albert.sites.beta.gouv.fr/solutions/models/

9. Accessibilité

Vous pouvez nous aider à améliorer l'accessibilité du site en nous signalant les problèmes éventuels que vous rencontrez. Pour ce faire, envoyez-nous un courriel à albert.api@numerique.gouv.fr.

Si vous nous avez signalé un défaut d'accessibilité vous empêchant d'accéder à un contenu ou à un des services et que vous n'avez pas obtenu de réponse satisfaisante de notre part, vous pouvez :

- Écrire un message au Défenseur des droits à l'aide du formulaire accessible sur https://formulaire.defenseurdesdroits.fr
- Envoyer un courrier par La Poste (gratuit, sans affranchissement) à :
  Défenseur des droits
  Libre réponse 71120
  75342 Paris CEDEX 07