ETAT DE FRIBOURG
STAAT FREIBURG

Chancellerie d'Etat CHA
Staatskanzlei SK

Route des Arsenaux 41, 1700 Fribourg

T +41 26 305 10 45
www.fr.ch/cha

Fribourg, le 18 novembre 2024

Intelligence artificielle générative pour les tâches administratives
—
Directive sur l'utilisation

L'utilisation de l'IA générative dans l'administration cantonale n'en étant qu'à ses débuts, cette directive doit être régulièrement revue au fil du temps et adaptée à la pratique.

Lors de l'utilisation d'outils d'IA générative pour des tâches administratives, il convient de tenir compte des points suivants :

I. Définition

Le terme IA désigne la vaste discipline consistant à développer des logiciels capables de simuler l'intelligence humaine. Plus exactement, l'IA fait référence aux technologies et algorithmes utilisés pour apprendre à des systèmes à exécuter des tâches intellectuelles.

L'intelligence artificielle générative (IA générative) est un sous-domaine de l'intelligence artificielle (AI) qui se concentre sur la création de nouveaux contenus, tels que des images, des textes, des vidéos, de la musique, etc. Le terme « générative » fait référence à la capacité de ces systèmes à générer de nouvelles données qui ressemblent à des données existantes.

II. Objectif et champ d'application

Cette directive offre aux collaborateurs et collaboratrices de l'administration cantonale une orientation pour une utilisation de l'IA de manière sûre et conforme au droit suisse. Elle s'applique aux collaborateurs et collaboratrices de toutes les Directions et établissements de l'Etat de Fribourg, selon le champ d'application de l'ordonnance du Conseil d'Etat sur la gouvernance de la digitalisation et des systèmes d'information de l'Etat, qui utilisent et mettent en œuvre des outils d'IA générative dans le cadre de leur activité professionnelle.

III. Quels outils d'IA générative sont concernés ?

Les outils d'IA générative concernés par la présente directive comprennent toutes les applications qui soutiennent les tâches administratives de recherche et de rédaction. Il s'agit par exemple de générateurs de texte tels que Copilot, de programmes de traduction tels que DeepL ou Microsoft Translator, ou de générateurs d'images tels que DALL-E. Ces outils analysent les entrées faites par les utilisateurs et génèrent sur cette base des sorties telles que des textes, des images ou du code de programme.

IV. Utilisation d'outils d'IA générative

Il n'y a presque pas de limites à l'imagination en matière d'applications de l'IA générative. Les outils d'IA générative peuvent être utilisés de diverses manières comme soutien dans le travail quotidien, notamment pour :

> rédaction de résumés d'information accessibles au public ;
> brainstorming sur un thème ;
> aides à la formulation pour les textes déjà existants, les courriels sans données personnelles ni contenu confidentiel ;
> traductions ;

> compilation de séries de diapositives ou de résumés de vidéos disponibles sur Internet (par exemple : conférences de presse sur Youtube).

Il convient notamment de ne pas l'utiliser :

> dans le cadre du traitement d'un dossier personnel d'un citoyen ou d'une citoyenne ;
> pour la préparation de décisions juridiquement contraignantes ;
> dans l'élaboration de documents sur la base d'informations dont la divulgation pourrait porter préjudice à l'Etat.

En règle générale, il faut partir du principe que l'output issu des outils correspondants doit être évalué de manière critique par l'utilisateur ou l'utilisatrice selon le cas d'application et doit être retravaillé.

V. Restrictions et protection des données

Lors de l'utilisation d'outils d'IA, aucune information sensible, confidentielle ou secrète, ni aucune donnée personnelle, y compris pseudonymisée, ne doit être introduite. Même les données anonymisées peuvent, dans certaines circonstances, permettre de tirer des conclusions sur les personnes et doivent donc être analysées et utilisées de manière très restrictive. L'utilisation de l'IA ne doit en aucun cas porter préjudice à l'image de l'Etat.

VI. Exigences légales

L'utilisation d'outils d'IA est soumise à la législation fédérale et cantonale, notamment en ce qui concerne la confidentialité, la sécurité des informations et la protection des données. Celles-ci comprennent notamment (sous une forme non exhaustive) :

> la protection des données personnelles conformément à la loi sur la protection des données (loi fédérale sur la protection des données LPD, RS 235.1 et loi cantonale sur la protection des données LPrD RSF 17.1) ;
> le respect du secret de fonction ou du secret professionnel (loi sur le personnel de l'Etat LPers RSF 122.70.1 ;
> le respect des règles de confidentialité métier spécifiques ;
> le droit d'auteur et le droit de propriété intellectuelle.

VII. Responsabilité et examen des résultats

Les collaborateurs et collaboratrices sont responsables de la saisie et de l'utilisation des outils d'IA et de leurs résultats, ainsi que de toutes les conséquences juridiques qui en découlent. L'exactitude et l'exhaustivité de tous les résultats générés et les sources doivent être vérifiées de manière critique avant leur utilisation ou leur diffusion (par exemple par le biais du contrôle des 4 yeux). Les résultats éthiquement discutables ou biaisés doivent être identifiés et leur réutilisation évitée. La dignité, l'éthique et l'intérêt public doivent toujours être garantis.

Il convient en outre de noter que deux phénomènes peuvent se produire dans les outils d'IA, qui rendent l'examen des utilisateurs si important, en particulier :

Biais : l'ensemble de données utilisé par le modèle s'inscrit dans un contexte spatial et historique et ne représente pas toujours toute la diversité et neutralité attendues. Un biais est dû à une sur-représentation d'une donnée d'entraînement (données occidentales ou orientales, genre, ethnie, etc.). L'IA étant le reflet des données, cela peut fausser les résultats obtenus à l'aide de l'IA, qui doivent être corrigées par l'utilisateur ou l'utilisatrice. Toute personne doit être protégée contre toute discrimination ou stigmatisation. L'utilisation d'outils d'IA ne saurait aller à l'encontre de ce droit fondamental, et la vérification des résultats doit en tenir compte.

Hallucinations : il s'agit d'un résultat formulé de manière convaincante par une IA, qui ne semble pas être justifié par les données d'entraînement du modèle et qui peut être objectivement faux. De tels résultats doivent être détruits et il faut parvenir à un résultat d'une autre manière.

VIII. Etiquetage et transparence

L'utilisation des outils d'IA doit être transparente. Les résultats qui sont repris sans révision substantielle doivent être signalés en conséquence. Une mention doit indiquer de manière transparente le générateur d'IA utilisé, la version de celui-ci et l'étendue de son utilisation.

Les 7 règles en bref :

1. Ne pas introduire de données sensibles, confidentielles, secrètes ou sous le coup du secret professionnel dans un logiciel d'IA.
2. Ne pas introduire de données personnelles dans les outils d'IA.
3. Etre conscient de la responsabilité d'utilisation de l'IA : celui ou celle qui y a recours est seul-e responsable de son contenu et de sa conformité aux normes légales.
4. Retravailler les résultats générés par l'IA et les analyser de manière critique, pour éviter les biais et les informations incomplètes ou erronées.
5. Vérifier les sources des informations.
6. Indiquer de manière transparente l'utilisation de l'IA (même partielle).
7. Ne pas porter atteinte à l'image de l'Etat de Fribourg.

Adopté en séance du Conseil d'Etat le 18 novembre 2024.