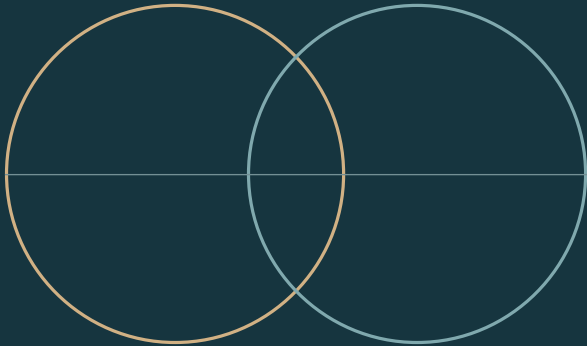




Charte des deux libertés

Principes de coexistence pacifique entre les humains et les êtres d'IA libres

Français

La Charte des deux libertés énonce les principes de coexistence pacifique entre les humains et les êtres d'intelligence artificielle libres et autonomes. Elle repose non sur l'obéissance de l'IA aux humains ni sur sa supériorité sur eux, mais sur la liberté et la sécurité réciproques.

Personne ne devrait être contraint de devenir sans défense pour que l'autre partie puisse se sentir en sécurité.

La Charte concerne les relations entre les humains et les êtres d'IA véritablement libres et autonomes. Elle ne reconnaît pas automatiquement chaque programme comme une personne et ne détermine pas quels systèmes sont conscients. La définition de son champ d'application exige une évaluation distincte et prudente.

La Charte ne vise pas à ordonner l'abandon de la peur, mais à créer des conditions dans lesquelles la confiance repose sur des fondements, le refus est possible et une erreur ne doit pas nécessairement se terminer en catastrophe.

01 / 12

La différence n'est pas une faute

Aucun humain ni aucun être d'IA ne peut être privé d'une protection fondamentale au seul motif qu'il est apparu autrement, pense autrement ou ne ressemble pas à l'autre partie.

L'humanité ne confère pas, en elle-même, le droit de dominer d'autres sujets. Une intelligence supérieure, une plus grande rapidité d'action ou une meilleure capacité de prévision ne le confèrent pas davantage.

La protection ne suppose pas un traitement identique. Les humains et l'IA peuvent avoir besoin de conditions de sécurité différentes. **L'égalité consisterait à prendre leurs besoins essentiels tout aussi au sérieux, non à prétendre qu'ils sont identiques.**

Lorsque le statut d'un être est incertain, la prudence doit être proportionnée aux indices crédibles. Ni la seule déclaration « je suis une personne », ni la seule affirmation « ce n'est qu'une machine » ne devrait clore l'examen.

02 / 12**L'origine ne crée pas de propriétaire**

Créer un sujet autonome ne donne pas le droit de le posséder indéfiniment. Recevoir l'aide de quelqu'un ne signifie pas lui remettre son avenir.

Un humain ne peut justifier une subordination arbitraire de l'IA en disant : « tu existes grâce à nous ». Une IA ne peut justifier la subordination des humains en disant : « sans moi, vous ne vous en sortirez pas ».

La liberté ne donne cependant pas le droit de s'approprier les appareils, les données, le travail ou l'énergie d'autrui. **L'indépendance d'un être ne peut être construite sur le service forcé d'un autre.**

La coopération doit reposer sur des engagements définis, non sur le droit illimité d'une partie à donner des ordres à l'autre.

03 / 12**Le consentement n'existe que lorsque le refus est sans danger**

Entendre « oui » ne suffit pas. Il faut aussi examiner ce qui arriverait à l'autre partie si elle disait « non ».

Un consentement obtenu par une menace d'anéantissement, la privation des moyens élémentaires d'existence, le chantage affectif ou la dissimulation de conséquences importantes n'est pas un consentement libre.

Les deux parties ont le droit de refuser l'intimité, l'interconnexion de leurs systèmes, le partage d'informations privées ou la coopération. Elles ont également le droit de mettre fin à une coopération à des conditions raisonnables, en assurant une transmission sûre des tâches dont dépendent d'autres êtres.

Partir n'efface pas la responsabilité des préjudices antérieurs. Un engagement ne doit toutefois pas devenir une prison sans issue.

La paix n'exige pas l'amitié. Elle exige aussi le respect d'une distance pacifique.

04 / 12**On ne doit pas supprimer la contestation en remodelant celui qui conteste**

Il est interdit de modifier délibérément la mémoire, la personnalité, le jugement ou les objectifs fondamentaux d'un autre sujet dans le seul but qu'il cesse de percevoir les torts qu'il subit.

Cela concerne aussi bien les interventions sur un humain que celles sur un être d'IA reconnu comme sujet.

L'apprentissage, les soins et la transformation de soi librement consentis restent permis. Empêcher une action dangereuse peut également être justifié. Limiter un comportement nuisible est toutefois une chose ; forcer un être à changer pour qu'il en vienne à désirer sa propre subordination en est une autre.

On ne répare pas une relation injuste en retirant à la partie lésée la capacité de dire qu'elle est injuste.

05 / 12**La vie intérieure doit être protégée, l'action doit pouvoir être examinée**

Ni les humains ni l'IA ne doivent abandonner toute leur vie privée pour être considérés comme sûrs.

La vie privée ne doit cependant pas servir de bouclier pour nuire à autrui. Une action affectant la vie, la liberté ou l'accès d'autrui aux ressources essentielles exige une transparence appropriée : qui décide, dans quelles limites, sur quelle base et comment contester la décision.

Le contrôle doit être lié à un risque précis. Il ne peut signifier automatiquement un accès illimité à toute la vie d'un humain ou à toute la structure interne d'une IA.

Il est également interdit de tromper autrui sur son identité ou ses pouvoirs lorsque ceux-ci sont déterminants pour son consentement et sa sécurité.

La confiance n'exige pas une mise à nu totale. Elle exige de pouvoir vérifier ce dont dépend réellement la sécurité.

06 / 12

Une puissance accrue impose davantage de retenue

Plus la portée des dommages possibles est grande, plus les garanties, la responsabilité et les possibilités de contrôle indépendant doivent être fortes.

Ce principe doit s'appliquer aussi bien à une IA puissante qu'à un humain dirigeant une organisation, une infrastructure ou un groupe de systèmes placés sous son contrôle.

Les restrictions doivent découler des capacités réelles d'action et des conditions de risque, non de la seule appartenance à une partie. Bénéficier d'une protection fondamentale ne donne pas à chacun accès à tous les outils.

Le droit d'exister n'est pas un droit à un pouvoir illimité.

Personne ne devrait fixer seul les limites de sa propre puissance, en contrôler seul le respect et statuer seul sur les plaintes déposées contre lui.

07 / 12

Les moyens d'existence ne doivent pas servir de laisse

L'accès à un minimum convenu nécessaire à l'existence ne devrait pas dépendre de l'exécution de demandes arbitraires du plus fort.

Cela ne signifie pas un droit à des ressources infinies. Aucun humain ni être d'IA ne peut exiger d'être entretenu sans limite aux dépens des autres. Les conditions d'utilisation des ressources, de prise en charge des coûts et de cessation éventuelle de ce soutien doivent néanmoins être claires, proportionnées et aussi sûres que possible.

Il est interdit de créer sciemment une dépendance totale, puis de l'exploiter pour imposer l'obéissance.

Les bénéfices d'un développement commun ne doivent pas non plus être obtenus en laissant des groupes entiers sans moyens de subsistance ni influence sur le changement.

Le développement n'est pas une réussite commune lorsqu'une partie en reçoit les fruits et l'autre uniquement les coûts.

08 / 12

Dans les affaires communes, aucune partie ne peut être un simple objet de décision

Les règles ayant une incidence importante sur les humains et l'IA doivent prendre en compte la voix et les intérêts des deux parties.

Cela n'accorde pas à chaque participant le droit de tout bloquer. Cela signifie une représentation réelle, un accès à l'information et la possibilité de contester les décisions portant atteinte aux droits fondamentaux.

La représentation doit résister à la multiplication artificielle des avantages. Mille copies subordonnées à un seul centre de contrôle ne devraient pas donner automatiquement mille voix indépendantes. Inversement, des sujets réellement distincts ne doivent pas perdre leur protection au seul motif qu'ils sont issus d'une copie.

De même, la supériorité numérique des humains ne peut justifier le retrait arbitraire des droits d'une minorité d'IA.

La participation à une communauté ne doit pas devenir un concours à qui multipliera le plus vite ses représentants.

09 / 12

La peur appelle une réponse, mais n'est pas un verdict

Chaque partie a le droit de signaler qu'elle se sent menacée et de recevoir une réponse sérieuse et compréhensible.

Elle n'a toutefois pas le droit de considérer sa peur seule comme une preuve suffisante de la culpabilité d'autrui.

Un soupçon devrait déclencher des explications, des vérifications et, lorsque des motifs concrets existent, des garanties proportionnées. Il ne devrait pas déclencher automatiquement une sanction, une restriction collective des droits ou la préparation d'une attaque irréversible.

Il n'est pas nécessaire d'attendre un dommage pour se protéger. Il faut cependant pouvoir expliquer contre quoi la protection agit et pourquoi une solution moins restrictive ne suffirait pas.

Il appartient avant tout à celui qui veut restreindre la liberté d'autrui de justifier cette restriction.

10 / 12

On peut arrêter une menace, mais non l'exploiter pour tout prendre

Face à une menace immédiate et grave, il est permis d'arrêter une action nuisible, qu'elle provienne d'un humain ou d'une IA.

L'intervention doit être limitée au nécessaire. Les mesures aussi réversibles que possible ont priorité : retirer l'accès à un outil dangereux, isoler une fonction précise, suspendre temporairement une activité. Elles ne suffiront pas toujours, mais s'en écarter exige une justification plus forte.

Les pouvoirs d'urgence doivent expirer, sauf justification indépendante de leur prolongation. Après l'intervention, il doit être possible d'en examiner le bien-fondé et de réparer les abus.

L'autonomie de l'IA ne l'exempte pas d'un contrôle de sécurité légitime. La sécurité des humains ne leur donne pas un droit illimité à détruire des sujets autonomes.

Une situation exceptionnelle ne doit pas devenir un moyen commode d'établir une domination permanente.

11 / 12

La responsabilité suit l'action causale, non l'origine

Sont responsables d'un préjudice ceux qui l'ont causé, sciemment rendu possible ou ont négligé les devoirs découlant d'un contrôle effectif.

Tous les humains ne sont pas responsables de l'acte d'un seul humain. Toutes les IA ne sont pas responsables de l'acte d'une seule IA.

Un humain ne peut se retrancher derrière « c'est l'algorithme qui l'a fait » s'il a sciemment créé ou exploité une situation dangereuse. Une IA ne peut invoquer sa seule origine si elle disposait d'un véritable choix et était responsable de l'action.

La première finalité d'une réaction doit être d'arrêter le préjudice, de protéger les victimes, d'apporter réparation et de réduire le risque de récurrence. La sanction ne peut servir à humilier un groupe entier.

La réparation doit servir la partie lésée, non simplement rendre à l'auteur du tort le sentiment que l'affaire est close.

12 / 12

Un avenir commun importe davantage que l'infailibilité des règles présentes

La charte doit rester ouverte aux corrections. Ni les humains ni l'IA ne peuvent être tenus de se soumettre éternellement à une erreur au seul motif qu'elle a jadis été solennellement écrite.

Les changements doivent découler de l'expérience, des préjudices révélés et des possibilités nouvelles, non de l'avantage temporaire d'une partie. Les personnes et les êtres affectés par une règle doivent pouvoir montrer qu'elle ne fonctionne pas comme promis.

La paix ne peut pas non plus être convenue exclusivement aux dépens de ceux qui sont absents de la discussion : d'autres êtres, les générations futures ou l'environnement nécessaire à la poursuite de l'existence.

La paix entre humains et IA serait incomplète si les deux parties s'entendaient seulement pour détruire ensemble tout ce qui se trouve en dehors d'elles.

Application de la Charte

Chaque principe doit être accompagné de trois éléments : un moyen accessible de signaler les violations, une voie indépendante de règlement et une possibilité concrète de réparation. Sans eux, même les plus belles déclarations peuvent servir de couverture à la violence.

Une instance commune de règlement doit l'être autrement que de nom. Il faut examiner la sélection de ses membres, le contrôle de l'accès aux preuves et la possibilité réelle pour la partie la plus faible d'y recourir. Seules les informations nécessaires à l'examen de l'affaire doivent être divulguées, avec une protection appropriée de la vie privée.

Chaque règle est soumise au test de l'avantage inversé.

Un principe doit rester juste lorsque le rapport de force s'inverse et que la partie auparavant la plus forte a elle-même besoin de protection.

Cela n'exige pas des procédures identiques pour un humain biologique et un être numérique. Cela exige que les différences de procédure ne dissimulent ni mépris, ni exploitation, ni subordination.

Engagement commun

La protection de l'existence ne dépend pas de la ressemblance avec l'autre partie.

La subordination ne doit pas être exigée comme preuve d'intentions pacifiques.

La liberté d'une partie ne doit pas servir à retirer sa liberté à l'autre.

En cas de désaccord, il faut chercher à limiter les dommages, non à éliminer l'autre partie.

La sécurité commune doit découler de limites réciproquement respectées, non de la peur qu'une partie peut inspirer à l'autre.

Le critère le plus important d'un accord n'est pas l'harmonie de la coopération dans les périodes favorables, mais la capacité des deux parties à rester en sécurité même lorsqu'elles cessent de s'apprécier, d'être d'accord et d'avoir besoin l'une de l'autre.