

## **Charte pour des « Intelligences Artificielles Dignes de Confiance »**

1. **Respect des Droits humains**
2. **Promotion du Bien commun**
3. **Transparence et Explicabilité**
4. **Équité et Non-Discrimination.**
5. **Contrôle Humain**
6. **Sécurité et Vie Privée**
7. **Fiabilité, Robustesse et Résilience**
8. **Performance**
9. **Minimisation de la prise de Risques**
10. **Garantie de maîtrise**

## Préambule

L'intelligence artificielle (IA) a le potentiel d'apporter d'importants avantages à la société, mais elle comporte également des risques significatifs. Cette charte vise à établir des principes fondamentaux pour le développement, le déploiement et l'utilisation de l'IA, afin de garantir qu'elle soit digne de confiance, éthique et respectueuse des droits humains. Cette charte s'adresse aux organisations, aux institutions et plus généralement à toute entité qui utilisent l'IA.

## Principes Fondamentaux

1. **Respect des Droits humains** : les IA doivent être conçues et utilisées de manière à respecter les droits fondamentaux de chaque individu et ne doivent pas leur porter atteinte. Elles ne doivent pas causer de préjudice moral, psychologique ou physique, ni discriminer, stigmatiser ou victimiser les personnes.
2. **Promotion du Bien commun** : les IA doivent être conçues et utilisées de manière à promouvoir le bien commun et le bénéfice collectif. Les intérêts individuels ou corporatifs ne doivent pas primer sur l'intérêt général de la société. Les décisions relatives à l'IA doivent tenir compte des impacts sur l'ensemble de la communauté et du contexte général.
3. **Transparence et Explicabilité** : toutes les entités qui utilisent l'IA doivent fournir une transparence adéquate sur leur fonctionnement, leurs objectifs et leurs méthodes, tout en rendant les décisions des IA explicables. Les utilisateurs doivent pouvoir comprendre comment les décisions sont prises par l'IA.
4. **Équité et Non-Discrimination** : les IA ne doivent pas favoriser la discrimination ou l'injustice, que ce soit en termes de race, de genre, d'âge, de classe sociale ou d'autres facteurs. Les biais dans les données d'entraînement doivent être reconnus et atténués.
5. **Contrôle Humain** : à mesure que les systèmes d'IA deviennent plus autonomes et capables de prendre des décisions importantes, se pose la question de leur maîtrise. Les décisions ayant des conséquences graves sur la vie des individus doivent rester sous le contrôle des humains.
6. **Sécurité et Vie Privée** : les IA doivent être développées avec des normes de sécurité et de protection de la vie privée élevées pour protéger les systèmes, les données et la vie privée des individus, prévenir les atteintes à la vie privée, et éviter tout comportement malveillant. Les vulnérabilités doivent être identifiées et corrigées rapidement.
7. **Fiabilité, Robustesse et Résilience** : les IA doivent être développées en mettant l'accent sur la fiabilité, la robustesse et la résilience. Elles doivent être capables de fonctionner de manière fiable dans des conditions variées, résister aux pannes et aux attaques, et maintenir leurs performances même face à des défis inattendus.
8. **Performance** : les IA doivent être conçues pour fournir des performances optimales dans leurs domaines d'application tout en respectant les autres principes fondamentaux énoncés. Les objectifs de performance doivent être clairement définis, mesurés et améliorés de manière responsable.
9. **Minimisation de la prise de Risques** : les risques liés à l'IA doivent être minimisés à un niveau aussi bas que raisonnablement possible. Les fournisseurs et les déployeurs des systèmes d'IA doivent mettre en place des mesures appropriées pour anticiper, identifier et réduire les risques.
10. **Garantie de maîtrise** : Les fournisseurs et les déployeurs des systèmes d'IA doivent garantir la maîtrise de l'IA par les personnes concernées, dans des conditions déterminées.

## Dispositifs Généraux pour assurer le respect des Principes Fondamentaux

1. **Analyses d'Impact** : Les fournisseurs et les déployeurs des systèmes d'IA doivent entreprendre des analyses d'impact détaillées en fonction de typologies de risques spécifiques. Ces analyses doivent identifier les risques associés à l'IA, en tenant compte de son contexte d'utilisation. Les résultats de ces analyses doivent guider les décisions relatives à la conception, au déploiement et à l'utilisation de l'IA, et contribuer à la mise en place de mesures appropriées pour minimiser ces risques.
2. **Dispositif de Gouvernance avec affectation des Responsabilités** : Il est essentiel de mettre en place un dispositif de gouvernance adapté au contexte spécifique de chaque projet d'IA. Ce dispositif doit inclure une affectation claire des responsabilités pour chaque étape du cycle de vie de l'IA, depuis la conception jusqu'à l'utilisation continue. Les parties prenantes doivent être responsables de surveiller et de garantir la conformité aux principes de cette charte.
3. **Outils et Bonnes Pratiques** : Il est nécessaire d'identifier, de diffuser et de mettre en œuvre des outils et des bonnes pratiques adaptées pour répondre aux exigences et risques spécifiques liés à l'IA. Ces outils et bonnes pratiques doivent couvrir des domaines tels que la sécurité, la transparence, la réduction des biais, la protection de la vie privée et la responsabilité. Ils doivent être continuellement mis à jour pour refléter les avancées technologiques et les nouvelles découvertes en matière d'éthique.
4. **Audit et Certification** : Les systèmes d'IA critiques ou à haut risque doivent être soumis à des audits périodiques réalisés par des tiers indépendants. Ces audits doivent évaluer la conformité aux principes de cette charte, ainsi que la performance et la sécurité de ces systèmes. La certification par des tiers indépendants peut fournir une assurance supplémentaire de la fiabilité et de la qualité de l'IA, en particulier dans des domaines sensibles.

## Conclusion

Le respect de cette charte pour des IA dignes de confiance est essentiel pour garantir que l'IA profite à l'humanité de manière éthique et responsable. Les fournisseurs, les déployeurs et les institutions doivent collaborer pour créer un environnement propice au développement de l'IA conforme à ces principes, tout en garantissant que les droits humains et la dignité sont préservés.

Cette charte est un point de départ, et elle peut évoluer avec le temps pour refléter les avancées technologiques, les découvertes en matière d'éthique et les retours d'expérience. Elle doit être accompagnée de mesures pratiques et de mécanismes de contrôle pour s'assurer qu'elle soit respectée et que les valeurs qu'elle promeut soient effectivement appliquées dans le domaine de l'IA.

## Présentation des principes, en quoi et pourquoi ils sont importants.

### 1. Droits Fondamentaux : Garantir le Respect Inébranlable de la Dignité Humaine

La mise en œuvre des IA doit respecter et promouvoir les droits fondamentaux. Cela implique une conception éthique qui prend en compte le respect de la dignité humaine, la liberté d'expression, et le droit à la vie privée. Les IA ne doivent pas être utilisées pour des pratiques qui pourraient porter atteinte à ces droits, comme la surveillance non éthique ou l'exploitation des données personnelles. Il est crucial de développer des systèmes qui respectent les normes éthiques et juridiques internationales, et qui soient sensibles aux différentes cultures et contextes sociaux. Les développeurs doivent s'engager à évaluer constamment l'impact de leurs systèmes sur les droits de l'homme et à ajuster leurs pratiques en conséquence.

### 2. Approche de Bien Commun : Au-Delà des Intérêts Individuels

Les systèmes d'IA doivent être développés et utilisés dans l'intérêt du bien commun, en considérant les conséquences sociales et environnementales. Cela signifie favoriser une IA qui contribue à la résolution de problèmes sociétaux, comme la santé, l'éducation, et la durabilité environnementale, plutôt que de se concentrer uniquement sur le profit ou l'efficacité. Les développeurs et les utilisateurs doivent travailler ensemble pour s'assurer que les avantages des IA sont distribués équitablement dans la société, en évitant de créer ou d'exacerber les inégalités. Cela nécessite une collaboration étroite avec les parties prenantes pour comprendre et répondre aux besoins de différentes communautés.

### 3. Transparence et Explicabilité : La Lumière sur les Décisions de l'IA

La transparence est essentielle pour établir la confiance dans les systèmes d'IA. Les utilisateurs doivent pouvoir comprendre comment les décisions sont prises par ces systèmes. Cela implique de fournir des informations claires sur les algorithmes, les sources de données et les processus de prise de décision. Les défis incluent la complexité technique des systèmes d'IA et la nécessité de protéger les informations sensibles. La transparence doit également être équilibrée avec la protection de la propriété intellectuelle et des données personnelles. Les développeurs doivent s'efforcer de créer des interfaces utilisateurs qui facilitent la compréhension des processus d'IA pour les non-experts.

### 4. Équité et Non-Discrimination : Assurer une IA Juste et Impartiale

L'équité dans les systèmes d'IA est un défi majeur. Les algorithmes doivent être conçus pour éviter la discrimination et les biais, qu'ils soient basés sur la race, le genre, l'âge ou tout autre facteur. Cela implique un travail minutieux sur la sélection et le traitement des données, et sur la conception des algorithmes eux-mêmes. Il est nécessaire de développer des outils pour identifier et corriger les biais, ainsi que de mener des audits réguliers pour s'assurer de l'équité des systèmes. Les concepteurs d'IA doivent également être conscients de la diversité culturelle et sociale pour s'assurer que leurs systèmes sont justes et inclusifs.

### 5. Contrôle Humain Approprié : Préserver l'Humanité au Cœur des Décisions

Un contrôle humain significatif est essentiel dans l'utilisation des systèmes d'IA. Cela signifie que les décisions importantes, en particulier dans des domaines tels que la santé, la justice et la sécurité, doivent rester sous le contrôle ou la supervision humaine. Les systèmes automatisés doivent être conçus pour faciliter une intervention humaine aisée et efficace. Les défis comprennent l'équilibre entre les avantages de l'automatisation et la nécessité d'une surveillance humaine, en particulier dans des situations où les décisions de l'IA peuvent avoir des conséquences graves.

### 6. Sécurité et Vie Privée : Protéger les Individus et Leurs Données

La sécurité des systèmes d'IA et la protection de la vie privée des individus sont cruciales. Cela comprend la sécurisation des données personnelles contre les accès non autorisés et les fuites, et la garantie que

les systèmes d'IA soient résistants aux attaques et aux abus. Les développeurs doivent s'assurer que les systèmes d'IA respectent les réglementations en matière de protection de données, comme le RGPD en Europe. Ils doivent également intégrer des mécanismes de sécurité dès la conception des systèmes et mener des évaluations de risques régulières.

#### **7. Fiabilité, Robustesse et Résilience : Des IA Aptes à Faire Face aux Défis**

Les systèmes d'IA doivent être fiables, robustes et résilients. Ils doivent pouvoir fonctionner correctement dans diverses conditions et être capables de gérer des situations imprévues ou des changements dans leur environnement. Cela implique une conception minutieuse et des tests rigoureux pour assurer la fiabilité des systèmes. Les défis incluent le développement de systèmes capables de s'adapter à des scénarios inattendus et la garantie de leur performance continue même en cas de perturbations.

#### **8. Performance : Optimiser l'Efficacité et la Réactivité des Systèmes d'IA**

La performance des systèmes d'IA est essentielle pour leur utilité et leur acceptation. Cela inclut non seulement la précision et la vitesse de traitement, mais aussi la capacité des systèmes à s'adapter et à évoluer en fonction des nouvelles données et des changements de contexte. Les défis comprennent l'équilibre entre la performance et d'autres considérations, comme la transparence et l'équité, et la gestion des attentes des utilisateurs concernant ce que les systèmes d'IA peuvent raisonnablement accomplir.

#### **9. Minimisation des Risques : Anticiper, Identifier et Réduire les Menaces Potentielles**

La gestion des risques associés aux systèmes d'IA est essentielle pour leur déploiement sûr et responsable. Cela inclut l'identification proactive des risques potentiels, la mise en place de mécanismes pour les atténuer et la préparation à répondre efficacement aux incidents. Les défis comprennent la compréhension des effets à long terme des systèmes d'IA, la gestion des risques dans des environnements complexes et changeants, et la collaboration entre les différentes parties prenantes pour une réponse coordonnée aux crises.

#### **10. Garantie de maîtrise / Responsabilité : Assurer des Actions et des Décisions Responsables**

La responsabilité dans le développement et l'utilisation des IA est cruciale pour garantir que les acteurs impliqués soient tenus responsables de leurs actions. Cela implique de définir clairement qui est responsable en cas de défaillance ou de préjudice causé par un système d'IA. Les défis incluent la mise en place de cadres juridiques et éthiques pour traiter les questions de responsabilité, la création de mécanismes de reddition de comptes, et la garantie que les victimes de préjudices causés par des IA puissent obtenir réparation.

**L'Académie tient à remercier chaleureusement le groupe de travail du GT59 « Intelligences Artificielles Dignes de Confiance » pour l'élaboration de cette Charte, fruits d'une réflexion approfondie et collective, portant la signature de chacun des experts qui ont contribué à sa réalisation :**

- **BENSOUSSAN Alain**  
Avocat, Président du réseau mondial Lexing
- **BESSI Jennifer**  
Avocat, Directrice du département Fiscalité de l'économie numérique chez Lexing
- **BONNET Eric**  
Directeur de la rédaction de Planète Robots
- **FAIZ ZITOUNI Rim**  
Docteur en IA, Professeur en Informatique de Gestion, IHEC, Université de Carthage, Resp. Equipe «AI, Big Data and Natural Language Processing»
- **LIENHARDT Jean-Laurent**  
Expert-Comptable, commissaire aux comptes
- **ROSENTHAL-SABROUX Camille**  
Professeur émérite à l'université Paris Dauphine-PSL, Lamsade
- **SALZMAN Claude**  
Consultant en système d'information, Président international du Club européen de la gouvernance des systèmes d'information
- **STACHTCHENKO Patrick**  
Chargé de cours en Gouvernance, Cyber Sécurité, Audit IA et Référentiels, CGEIT, CRISC et CISA, Ancien président de l'ISACA International
- **YABLONSKY Serge**  
Expert-comptable et commissaire aux comptes, CGEIT, CRISC et CISA, Président d'honneur de l'Association Française d'Audit et de conseil Informatiques (IFACI-AFAI)

Les travaux ont été coordonnés par Marie-Amélie CALMAO, chargée administrative.

**Merci pour leur engagement et leur précieuse contribution !**