



Vous trouverez tous  
les documents ici :



[go.akademien-schweiz.ch/atable](https://go.akademien-schweiz.ch/atable)

**Suivi 4/24 • 3 décembre 2024**

## **Réglementer l'IA ? Perspectives pour la Suisse**

### **1. Constats**

#### **1.1. Résumé des exposés (diapositives disponibles en ligne)**

##### **L'IA de demain sera-t-elle plus fiable ? Ce que la recherche développe aujourd'hui**

*Philippe Cudré-Mauroux, professeur d'informatique à l'Université de Fribourg*

L'intelligence artificielle couvre une série de technologies différentes qui simulent l'intelligence humaine et peuvent se charger de tâches complexes. Actuellement, les modèles linguistiques géants tels que GPT4, qui ont été entraînés à l'aide de dizaines de milliers de cartes graphiques et d'une quantité absurde de données (ce qui nécessite des investissements à hauteur de centaines de millions de dollars), sont particulièrement populaires. Leurs résultats sont impressionnants : ces gigantesques modèles ont été testés dans de nombreux domaines et fournissent souvent des réponses meilleures que celles de la majorité des candidats humains.

Ces résultats sont malheureusement peu fiables : les modèles sont opaques et leurs résultats ne reflètent que des probabilités et non des faits. Ils reproduisent les caractéristiques des données d'apprentissage, et leurs résultats sont souvent décevants sur de nouveaux cas qui ne sont pas inclus dans les données d'apprentissage. Ces restrictions ne peuvent être évitées techniquement. La recherche travaille toutefois sur des approches qui détectent, amenuisent, ou contrebalancent ces problèmes. Par exemple, en rendant les résultats plus explicables, en intégrant mieux l'être humain dans la validation des résultats, ou en combinant différentes méthodes d'IA pour créer des systèmes plus robustes.

##### **Que signifie l'AI Act de l'UE pour la Suisse ?**


*Nadja Braun Binder, professeure de droit public à l'Université de Bâle*


Le règlement sur l'IA de l'UE vise à renforcer le marché intérieur et à promouvoir une IA axée sur l'humain et digne de confiance tout en protégeant la santé, la sécurité et les droits fondamentaux. Elle constitue un règlement relatif à la sécurité des produits et prévoit pour les systèmes IA à usage spécifique quatre niveaux de risque : pratiques interdites, haut risque, risque limité et risque minimal. Les systèmes IA à usage général et les modèles de base sont soumis à des obligations supplémentaires relatives à la transparence ainsi que d'autres obligations en cas de risque systémique.

De nombreuses dispositions du règlement sur l'IA sont vagues et doivent être concrétisées ; cela sera réalisé en partie dans les normes techniques qui deviendront de fait aussi pertinentes pour la Suisse. Une reprise du règlement sur l'IA pourrait engendrer des doubles emplois, n'offrirait aucune garantie en ce qui concerne la reconnaissance des procédures d'évaluation de la conformité et pourrait renforcer la position sur le marché des grandes entreprises technologiques.

**Académies suisses des sciences (a+) • Secrétariat général**

Maison des Académies • Laupenstrasse 7 • Case postale • 3001 Berne • Suisse

+41 31 306 92 20 • [info@akademien-schweiz.ch](mailto:info@akademien-schweiz.ch) • [akademien-schweiz.ch](https://www.akademien-schweiz.ch)  [@academies\\_ch](https://twitter.com/academies_ch)

 [swiss\\_academies](https://www.instagram.com/swiss_academies)

## **Saisir les opportunités de l'IA en gardant le contrôle - comment cela peut-il fonctionner ?**

*Thomas Burri, professeur de droit européen et de droit international public à l'Université de Saint-Gall*

L'AI Act de l'UE repose sur l'idée d'un contrôle humain de l'IA. Il contient notamment une réglementation relative à la surveillance humaine des utilisations de l'IA qui entrent dans la « catégorie à haut risque ». Mais est-ce vraiment possible et cela a-t-il un sens ? Thomas Burri fait la distinction entre IA « chaude » (p. ex. atterrissage d'un avion qui exige des réactions immédiates) et « froide » (p. ex. analyses médicales). Dans le cadre du programme national de recherche 77 « Transformation numérique », la question de savoir dans quelle mesure le contrôle humain est utile pour permettre à un objet volant d'atterrir en toute sécurité avec l'aide de l'IA (IA « chaude ») a été abordée. Résultat : les êtres humains ne sont pas particulièrement efficaces pour évaluer si l'IA est sur la bonne voie dans la résolution d'un problème ou non et, par conséquent, si une intervention de l'humain est pertinente. Les mesures prises pour rendre l'IA plus compréhensible n'ont pas nécessairement amélioré le taux de réussite. Généralement, on confie plus de responsabilités à une « équipe » associant l'IA et l'humain qu'à l'IA ou l'être humain seuls. Le risque existe que l'être humain devienne le bouc émissaire en cas d'erreurs commises par l'IA. Sur la base des résultats scientifiques, Thomas Burri a présenté huit pistes de réflexion, voir présentation.

### 1.2. Discussion

#### **Sur le plan politique, l'exigence de transparence est une évidence. Compte tenu de la thématique de la boîte noire, devons-nous l'abandonner ?**

En principe, nous pouvons savoir comment une IA fonctionne et comment elle a été entraînée. Toutefois, même dans ce cas, il est impossible pour les utilisateurs-trices de comprendre comment une IA arrive à une solution déterminée. La transparence n'est pas possible dans ce cas. En revanche, une déclaration est possible, c'est-à-dire que l'on a l'obligation de mentionner l'utilisation de l'IA. On peut aussi exiger de la transparence concernant les données utilisées pour l'entraînement.

La question de l'utilité de la transparence se pose également. Nous utilisons avec succès beaucoup de choses (p. ex. ordinateur) sans devoir comprendre leur fonctionnement exact. La transparence va aussi de pair avec le risque de rejeter la responsabilité, comme c'est le cas aujourd'hui pour les bannières de cookies. On submerge les utilisateurs-trices d'un nombre inacceptable d'informations complexes et on leur fait endosser la responsabilité de leur action. Il est préférable de rendre l'IA plus sûre ou de réfléchir comment nous pouvons éviter les plus grands dommages potentiels.

#### **L'AI Act de l'UE est un essai dans le but de trouver une approche intelligente. La question se pose pour les pays du monde entier. N'est-ce pas un coup d'épée dans l'eau ? L'évolution fulgurante nous est imposée par les grands acteurs. N'est-ce pas plutôt une question relevant du droit de la concurrence ?**

La question est parfaitement fondée. Toutefois, l'Etat a tout à fait des possibilités d'agir et les « Big Tech » sont tout à fait intéressées par des réglementations. Avec une approche pragmatique qui s'adapte aux évolutions, il est possible de limiter les risques. L'UE a aussi dû adapter l'AI Act parce que les grands modèles linguistiques ont fait leur apparition pendant son élaboration. De plus, il est important de promouvoir l'innovation pour avoir une multitude d'acteurs. Malgré l'excellence de la recherche dans les hautes écoles, il ne se passe pas grand-chose à ce niveau en Suisse, par exemple dans le développement de l'IA pour des utilisations spécifiques.

#### **Il semble judicieux que l'AI Act de l'UE soit fondé sur les risques. Toutefois, cela va de pair avec une obligation d'autorisation pour les approches risquées : cela peut-il fonctionner ? Ne devrait-on pas travailler plus clairement avec des interdictions ?**

Une certification représente un défi, mais elle est possible. De bonnes expériences ont été par exemple faites avec les normes techniques relatives à la cybersécurité. Elles s'appliquent à l'international et peuvent constituer un fondement pour une autorisation. Il est important d'avoir en

Suisse aussi un organisme de certification pour s'assurer d'avoir la compétence à l'intérieur de notre pays.

**L'approche de réglementer l'IA en premier lieu par le biais de diverses lois spéciales est convaincante. Existe-t-il une sorte de cartographie qui indique les domaines dans lesquels l'IA est aujourd'hui déjà suffisamment traitée ou les domaines dans lesquels il y a des lacunes ?**

Un aperçu de ce type serait utile, mais il n'existe malheureusement pas (encore). En principe, un grand nombre de lois doivent être examinées pour déterminer si elles réglementent l'IA de façon appropriée et s'il faut les adapter le cas échéant. Pour les thèmes les plus urgents, par exemple le trafic, cela se fait déjà. Dans le domaine de la médecine en revanche, la densité réglementaire est déjà suffisante ; ici, c'est plus la question d'harmonisations qui se pose. En plus d'aborder les thèmes les plus urgents, on devrait aussi s'attaquer aux lois transversales comme la loi sur la protection des données.

**La Suisse devrait-elle reprendre l'AI Act si l'accord-cadre aboutissait ?**

Non, presque certainement pas, étant donné qu'il n'y a pas d'accord bilatéral concernant l'AI Act. Vu que le projet du contrat ou des contrats à venir n'a pas encore été publié, il n'est pas possible de répondre à cette question de façon définitive.

**Avec la « Swiss AI Initiative », les EPF souhaite développer un propre approche. Est-ce une façon d'aborder les problèmes relatifs par exemple aux biais ou à la traçabilité ?**

L'avenir le dira. L'accès aux données est fondamental. Quelle est la valeur des données ? Comment obtient-on, par exemple l'Etat, des données de bonne qualité, y compris les métadonnées ? Les métadonnées nous permettent de trouver aussi une solution pour démontrer la qualité des données. Raison pour laquelle l'UE réglemente aussi l'accès aux données, il s'agit d'une approche importante. La Suisse a les compétences pour jouer un rôle de pionnier.

Le problème des biais dans les données ne sera pas simple à appréhender par la réglementation. Les biais qui se renforcent eux-mêmes sont particulièrement problématiques. Exemple : aux Etats-Unis, les personnes de couleur font moins confiance au système de santé, raison pour laquelle il y a moins de données. Et parce qu'il y a moins de données, les personnes de couleur sont moins bien traitées, donc la confiance continue de baisser.

**En plus des « risques d'exploitation », l'IA amène aussi des risques sociaux. Pouvez-vous développer ?**

Oui, les bouleversements sont potentiellement importants et rapides. Jusqu'il y a peu de temps, beaucoup disaient à leurs enfants ou petits-enfants que programmeur·euse était un emploi sûr. Est-ce encore le cas ? Des secteurs d'activités prétendument très sûrs sont mis sous pression par l'automatisation. Dans le domaine des traductions, il existe déjà depuis dix ans des logiciels basés sur l'IA et un grand nombre de traducteurs·trices perdent énormément de mandats. Qu'est-ce que cela implique pour la formation et la formation continue si les machines peuvent assumer des tâches de plus en plus complexes ?

## 2. Approfondissement

Vous trouverez l'ensemble des documents complémentaires en ligne sur [go.akademien-schweiz.ch/atable](https://go.akademien-schweiz.ch/atable) :

- Présentations
- Contacts scientifiques
- Policy Paper « Approche humaine à l'intelligence artificielle et sa réglementation » de Thomas Burri, Markus Christen et d'autres  
[https://www.unisg.ch/fileadmin/user\\_upload/HSG\\_ROOT/\\_Kernauftritt\\_HSG/Universitaet/Schools/LAW/Faculty/Burri/241118\\_MAIC\\_screen\\_fr.pdf](https://www.unisg.ch/fileadmin/user_upload/HSG_ROOT/_Kernauftritt_HSG/Universitaet/Schools/LAW/Faculty/Burri/241118_MAIC_screen_fr.pdf)
- Diverses publications sur l'IA de TA-SWISS, e.a. sur les deepfakes ; sur la reconnaissance automatique de la voix, de la langue et du visage ou sur ChatGPT

<https://www.ta-swiss.ch/fr/sujets#societe-de-linformation>

Une mise au concours sur les grands modèles de langage ainsi qu'une étude sur l'utilisation des données de la santé sont en cours à TA-SWISS.

- Divers articles de blog et d'autres contributions d'expert·e·s au sujet du droit et de la technique ainsi que des données de l'IA et l'IA en général :

<https://www.satw.ch/fr/intelligence-artificielle>