Égalité, intelligence artificielle, algorithmes et décisions automatisées : comment détecter et traiter la discrimination dans le contexte de l'IA ?

Fabian Lütz

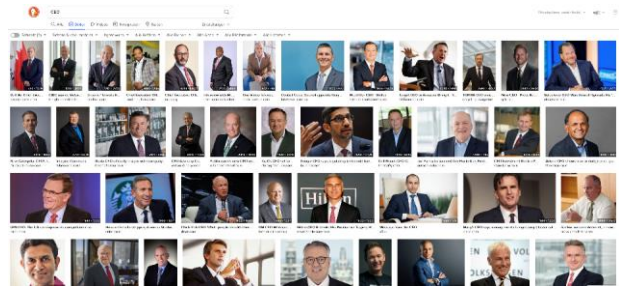
ERA (Trèves 12-13 juin 2023)

Financé dans le cadre du programme « Droits, égalité et citoyenneté » 2021-2027 de la Commission européenne

1

L'égalité des sexes à l'ère des algorithmes

Google (10.10.2021)



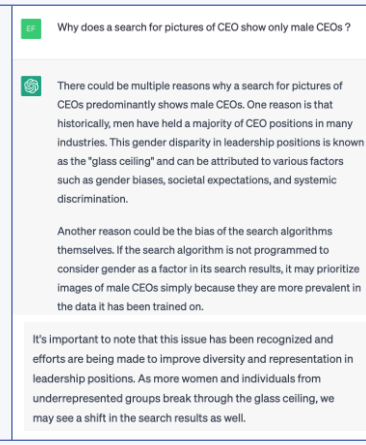
DuckDuckGo "CEO" (10/10/2021)

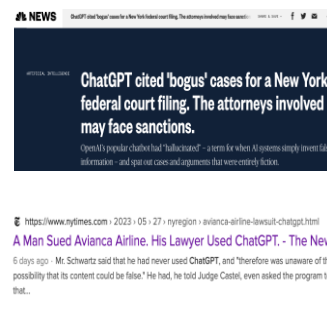
2

Grands modèles linguistiques (GMLs) : ChatGPT & Co

Le problème des préjugés sexistes et des discriminations potentielles peut augmenter avec l'utilisation de ChatGPT, GPT-4, Bard, etc.

→ Les modèles d'IA ou de fondation à usage général ne sont pas inclus dans la proposition COM (mais le Conseil a demandé que la GPAI soit prise en compte).



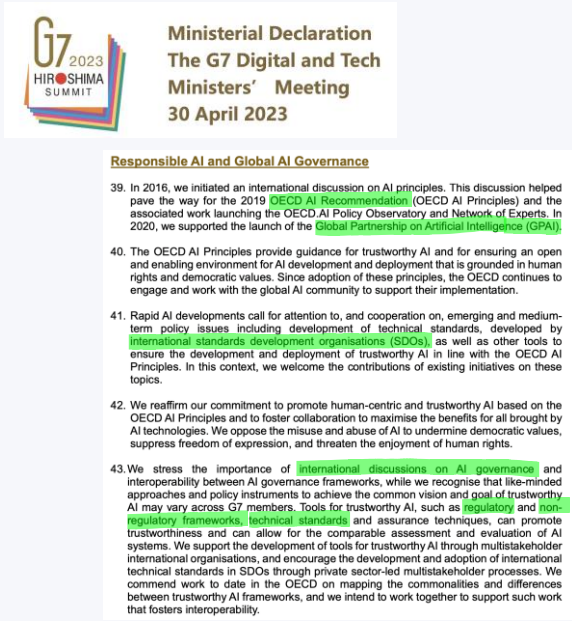


3

Ordre du jour

1. *Introduction*
2. *Cadre juridique actuel et futur de l'UE*
3. *L'IA et l'égalité des sexes en pratique*
4. *Réglementation de l'IA et des algorithmes*
5. *L'IA, catalyseur de l'égalité entre les hommes et les femmes ?*
6. *Perspectives*

4



Ministerial Declaration
The G7 Digital and Tech Ministers' Meeting
30 April 2023

Responsible AI and Global AI Governance

39. In 2016, we initiated an international discussion on AI principles. This discussion helped pave the way for the 2019 OECD AI Recommendation (OECD AI Principles) and the associated work launching the OECD.AI Policy Observatory and Network of Experts. In 2020, we supported the launch of the Global Partnership on Artificial Intelligence (GPAI).

40. The OECD AI Principles provide guidance for trustworthy AI and for ensuring an open and enabling environment for AI development and deployment that is grounded in human rights and democratic values. Since adoption of these principles, the OECD continues to engage and work with the global AI community to support their implementation.

41. Rapid AI developments call for attention to, and cooperation on, emerging and medium-term policy issues including development of technical standards, developed by international standards development organisations (SDOs), as well as other tools to ensure the development and deployment of trustworthy AI in line with the OECD AI Principles. In this context, we welcome the contributions of existing initiatives on these topics.

42. We reaffirm our commitment to promote human-centric and trustworthy AI based on the OECD AI Principles and to foster collaboration to maximise the benefits for all brought by AI technologies. We oppose the misuse and abuse of AI to undermine democratic values, suppress freedom of expression, and threaten the enjoyment of human rights.

43. We stress the importance of international discussions on AI governance and interoperability between AI governance frameworks, while we recognise that like-minded approaches and policy instruments to achieve the common vision and goal of trustworthy AI may vary across G7 members. Tools for trustworthy AI, such as regulatory and non-regulatory frameworks, technical standards and assurance techniques, can promote trustworthiness and can allow for the comparable assessment and evaluation of AI systems. We support the development of tools for trustworthy AI through multistakeholder international organisations, and encourage the development and adoption of international technical standards in SDOs through private sector-led multistakeholder processes. We commend work to date in the OECD on mapping the commonalities and differences between trustworthy AI frameworks, and we intend to work together to support such work that fosters interoperability.

1. Introduction

- Définitions
- Déclaration de Toronto
- Rappel des problèmes liés à l'égalité entre les hommes et les femmes
- Mise à jour juin 2023

Définitions : IA et algorithmes

Intelligence artificielle

- Un système d'IA est un logiciel qui "(..) pour un ensemble donné d'objectifs définis par l'homme, génère des résultats tels que du contenu, des prédictions, des recommandations ou des décisions influençant les environnements avec lesquels ils interagissent ;(..)".

  (Art. 3 para. 1, Loi sur l'IA de l'UE)

- Exemples :
  - apprentissage automatique (supervisé, non supervisé, renforcement, apprentissage profond), approches statistiques, méthodes de recherche et d'optimisation.

Algorithme

- "... des instructions suffisamment détaillées et systématiques pour résoudre un problème mathématique de telle sorte que, lorsqu'elles sont appliquées correctement, l'ordinateur calcule la sortie correcte pour chaque ensemble correct d'entrées"

  (Prof. Katharina Zweig)

Définitions : Erreurs de décision

Biais

- Biais cognitifs, terme générique désignant les erreurs de décision systématiques

(Daniel Kahneman, Olivier Sibony, Cass R. Sunstein : Noise, 2021, p. 163f.)

Bruit

- Variabilité de l'erreur, variabilité aléatoire des jugements *ou* taxe invisible

(K/S/S : Noise, 2021, p. 3-5 et D. Kahneman, HBR, 2016, p. 36-43)

La discrimination algorithmique se produit lorsque des systèmes automatisés contribuent à une différence de traitement injustifiée ou à des effets défavorables sur les personnes, par exemple en fonction de leur sexe.

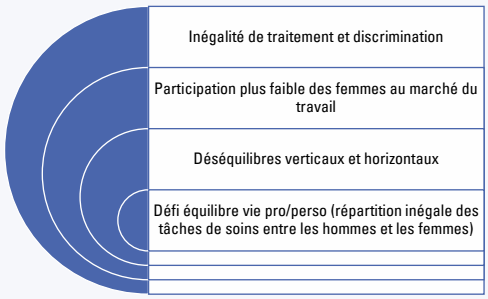
Source : https://www.torontodeclaration.org

Déclaration de Toronto

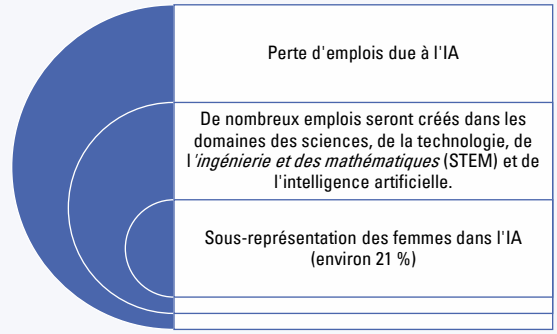
- Devoir des Etats de protéger les droits de l'homme ; cela inclut la garantie du droit à la non-discrimination par les acteurs du secteur privé" (Toronto, para. 38)
- Nécessité d'une surveillance de l'utilisation de l'apprentissage automatique par le secteur privé dans des contextes qui présentent un risque de résultats discriminatoires (Toronto, paragraphe 40).

Rappel de quelques problèmes liés à l'égalité entre les hommes et les femmes (ex. marché du travail)

Général



Spécifique à l'IA



9

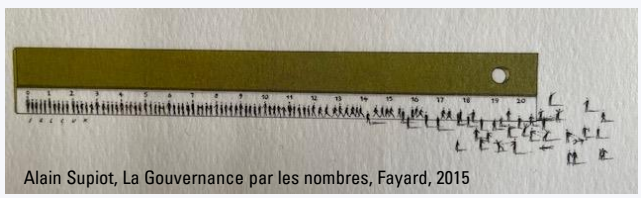
Dernières nouvelles de la mise à jour générale : Juin 2023

- Sommet RightsCon Costa Rica 5-8 juin 2023 : Les droits de l'homme à l'ère numérique
- Les technologies nouvelles et émergentes doivent faire l'objet d'une surveillance urgente et d'une transparence solide : Experts de l'ONU (2 juin 2023) → genre et discrimination
- Note d'information du Secrétaire général des Nations unies sur le Pacte mondial pour le numérique → approche multipartite (intéressant pour les entreprises et les conseillers juridiques)
- Consultation des Nations Unies sur les droits de l'homme & sur le genre, la technologie et le rôle des entreprises (15 juin 2023)

10

Dernières nouvelles de la mise à jour de l'UE : Juin 2023

- Texte de compromis du Parlement européen (PE) (16 mai 2023)
- Vote en plénière du Parlement européen prévu du 12 au 15 juin 2023
- *"Une fois approuvées, elles constitueront les premières règles mondiales en matière d'intelligence artificielle"* (EP News, 11 mai 2023).
- Les députés introduisent des plaintes concernant les systèmes d'IA + reçoivent des explications sur les décisions basées sur des systèmes d'IA à haut risque qui ont un impact significatif sur leurs droits.
- Renforcement de l'Office européen de l'IA chargé de contrôler la mise en œuvre du règlement sur l'IA

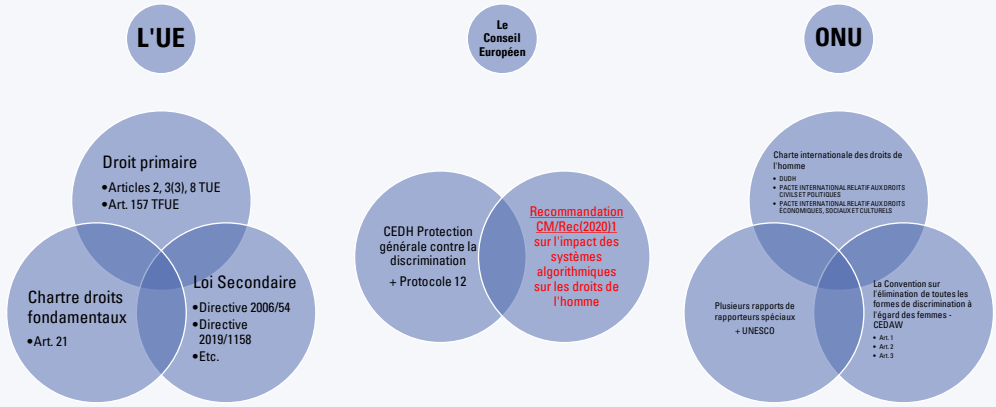


Alain Supiot, La Gouvernance par les nombres, Fayard, 2015

2. Cadres juridiques actuels et futurs

- Droit applicable
- Effet Bruxelles ?
- Proposition de l'UE : loi sur l'IA

Cadres juridiques actuels



Futur cadre juridique

Règlement de l'UE sur l'IA

- DSA
  - ECAT/JRC Sevilla
- DMA
- Loi sur l'IA
- Directive sur les normes EB
  - Récital 16
- Directive sur la responsabilité en matière d'IA
- Amélioration de la réglementation/révision de l'acquis communautaire + prochain mandat 2024-2029)

Conseil de l'Europe

- Généralités :
  - Convention-cadre du Conseil sur l'IA, les droits de l'homme, la démocratie et l'État de droit (adoption en novembre 2023 ?)
- GE/discrimination :
  - Instrument juridique pour lutter contre l'égalité des sexes et la discrimination (jusqu'en 2025)

ONU

- ONU
  - Pacte numérique mondial
- HCDH
  - Priorité, Rapports
- CEDAW
  - Recommandation générale 40
- CSW
  - Conclusions concertées "CSW 67

15

L'"effet Bruxelles" pour réglementer la discrimination algorithmique

L'UE à l'avant-garde de l'élaboration de la réglementation relative à l'IA et à la discrimination algorithmique

- Voir Anu Bradford (en général+RGPD)
- L'UE a ouvert un bureau à San Francisco pour suivre l'évolution de l'IA et de la technologie (les *États-Unis créent l'IA, l'UE réglemente l'IA*).
- Après l'adoption de la DSA/DMA + ECAT, l'UE sera probablement la première grande région économique à disposer d'une surveillance réglementaire de grande envergure pour les systèmes d'IA avec la loi sur l'IA.

PRESS RELEASE | Publication 05 September 2022

EU opens new office in San Francisco to reinforce its digital diplomacy

The European Union has opened its new office in San Francisco, California, a global centre for digital technology and innovation. The office will reinforce the EU's cooperation with the United States on digital diplomacy and strengthen the EU's capacity to reach out to key public and private stakeholders, including policy makers, the business community, and civil society in the digital technology sector.

High Representative/Vice-President Josep **Borrell**, said:

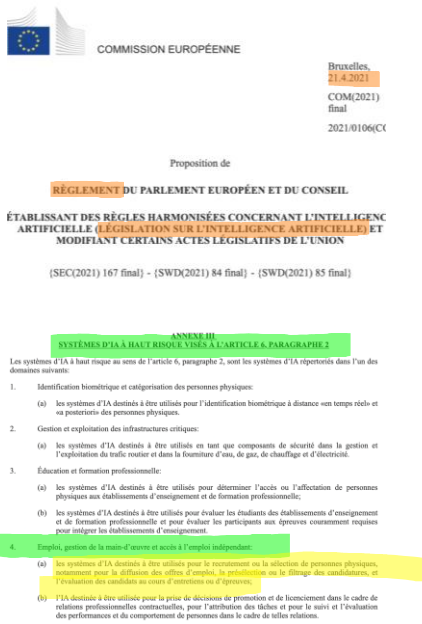
> The opening of the office in San Francisco responds to the EU's commitment to strengthen transatlantic technological cooperation and to drive the global digital transformation based on democratic values and standards. It is a concrete step to further reinforce the EU's work on issues such as cyber and countering hybrid threats, and foreign information manipulation and interference.

The EU office in San Francisco will seek to promote EU standards and technologies, digital policies and regulations and governance models, and to strengthen cooperation with US stakeholders, including by advancing the work of the EU-US Trade and Technology Council. Find more information in the EEAS press release.

Related topics

International relations

16



COMMISSION EUROPÉENNE

Bruxelles,
21.4.2021
COM(2021)
final
2021/0106(C

Proposition de

RÈGLEMENT DU PARLEMENT EUROPÉEN ET DU CONSEIL

ÉTABLISSANT DES RÈGLES HARMONISÉES CONCERNANT L'INTELLIGENC
ARTIFICIELLE (LÉGISLATION SUR L'INTELLIGENCE ARTIFICIELLE) ET
MODIFIANT CERTAINS ACTES LÉGISLATIFS DE L'UNION

{SEC(2021) 167 final} - {SWD(2021) 84 final} - {SWD(2021) 85 final}

ANNEXE III
SYSTÈMES D'IA À HAUT RISQUE VISÉS À L'ARTICLE 6, PARAGRAPHE 2

Les systèmes d'IA à haut risque au sens de l'article 6, paragraphe 2, sont les systèmes d'IA répertoriés dans l'un des domaines suivants:

1. Identification biométrique et catégorisation des personnes physiques:
   (a) les systèmes d'IA destinés à être utilisés pour l'identification biométrique à distance «en temps réel» et «a posteriori» des personnes physiques.
2. Gestion et exploitation des infrastructures critiques:
   (a) les systèmes d'IA destinés à être utilisés en tant que composants de sécurité dans la gestion et l'exploitation du trafic routier et dans la fourniture d'eau, de gaz, de chauffage et d'électricité.
3. Éducation et formation professionnelle:
   (a) les systèmes d'IA destinés à être utilisés pour déterminer l'accès ou l'affectation de personnes physiques aux établissements d'enseignement et de formation professionnelle;
   (b) les systèmes d'IA destinés à être utilisés pour évaluer les étudiants des établissements d'enseignement et de formation professionnelle et pour évaluer les participants aux épreuves couramment requises pour intégrer les établissements d'enseignement.
4. Emploi, gestion de la main-d'œuvre et accès à l'emploi indépendant:
   (a) les systèmes d'IA destinés à être utilisés pour le recrutement ou la sélection de personnes physiques, notamment pour la diffusion des offres d'emploi, la présélection ou le filtrage des candidatures, et l'évaluation des candidats au cours d'entretiens ou d'épreuves;
   (b) l'IA destinée à être utilisée pour la prise de décisions de promotion et de licenciement dans le cadre de relations professionnelles contractuelles, pour l'attribution des tâches et pour le suivi et l'évaluation des performances et du comportement de personnes dans le cadre de telles relations.

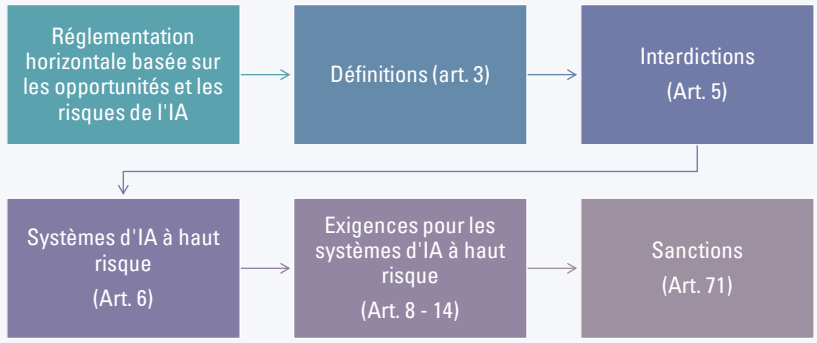
Loi européenne sur l'intelligence artificielle COM (2021), 206 final

Article 2
Champ d'application

1. le présent règlement s'applique
(a) les prestataires qui mettent sur le marché ou mettent en service des systèmes d'IA dans l'Union, qu'ils soient établis dans l'Union ou dans un pays tiers ;
(b) les utilisateurs de systèmes d'IA situés dans l'Union ;
(c) aux fournisseurs et aux utilisateurs de systèmes d'IA situés dans un pays tiers, lorsque les résultats produits par le système sont utilisés dans l'Union ;

17

UE COM(2021), 206 final Loi sur l'intelligence artificielle (LIA)



18

COM de l'UE Loi sur l'IA et l'égalité

Instrument horizontal pour l'IA

- La proposition complète de la législation de l'Union en vigueur en matière de non-discrimination
- Exigences concrètes pour minimiser le risque de discrimination par les algorithmes
- Conception et qualité des ensembles de données utilisés pour le développement des systèmes d'IA.
- Règles pour : les tests, la gestion des risques, la documentation et la surveillance humaine tout au long du cycle de vie des systèmes d'IA.

L'égalité dans la proposition de la loi IA

- Mention de la "non-discrimination" (16x)
- "Égalité entre les hommes et les femmes" (1x)
- "Femmes" (x2) : Art. 21 de la Charte des droits fondamentaux + dans les systèmes de recrutement

Projet de loi de l'UE sur l'IA et recrutement

Approche fondée sur les risques

- Article 6 + Article 6, paragraphe 2 + Annexe III (n° 4 Systèmes de recrutement et de sélection AI)
  - Pertinence pour l'égalité entre les femmes et les hommes
  - Danger pour les droits fondamentaux ?

Annexe III (Systèmes de recrutement)

- "Tout au long du processus de recrutement et lors de l'évaluation, de la promotion ou du maintien des personnes dans des relations contractuelles liées au travail, ces systèmes peuvent perpétuer des schémas historiques de discrimination, par exemple à l'égard des femmes (...)" (considérant 36).

Projet d'amendements du Parlement européen (16/05/2023)
Article 4 bis - Principes généraux applicables à tous les systèmes d'IA
1.e) *"**diversité, non-discrimination et équité**" signifie que les systèmes d'IA sont développés et utilisés de manière à inclure divers acteurs et à **promouvoir l'**égalité d'accès, l'**égalité des sexes** et la diversité culturelle, **tout en évitant les effets discriminatoires et les préjugés injustes qui sont interdits par le droit de l'Union ou le droit national**".*

COM de l'UE - Projet d'accord préalable en connaissance de cause et annexe III

Annexe III (Systèmes de recrutement), n° 4

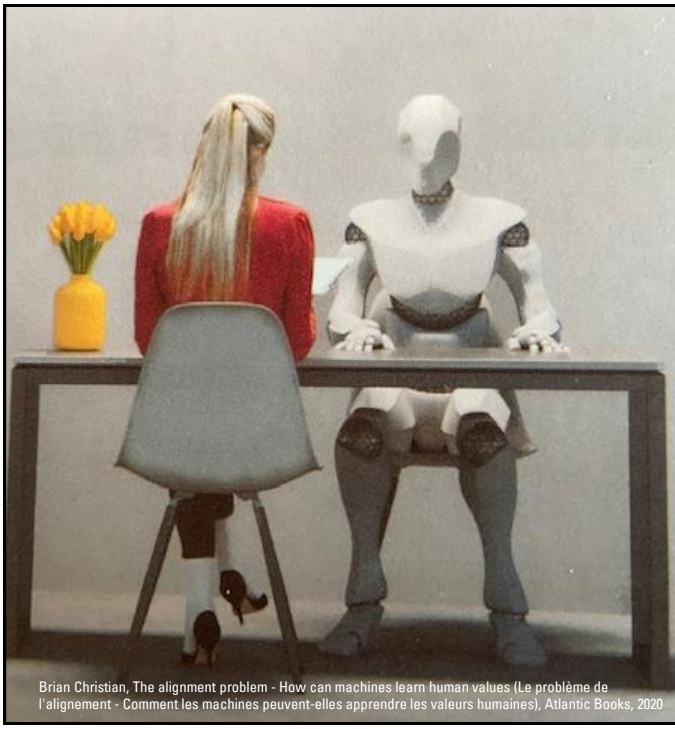
- L'emploi, la gestion des travailleurs et l'accès à l'auto-emploi :

(a) Les systèmes d'IA destinés à être utilisés pour le recrutement ou la sélection de personnes physiques, notamment pour la publication d'offres d'emploi, la sélection ou le filtrage des candidatures, l'évaluation des candidats au cours d'entretiens ou de tests ;

(b) L'IA destinée à être utilisée pour prendre des décisions sur la promotion et la résiliation des relations contractuelles liées au travail, pour la répartition des tâches et pour le suivi et l'évaluation des performances et du comportement des personnes dans le cadre de ces relations.

Article 6, paragraphe 2

- Classification en tant que système d'IA à haut risque
- Application des dispositions particulières du chapitre 2 (articles 8 à 14) :
  - Article 9 : Système de gestion des risques
  - Article 10 : données et gouvernance des données
  - Article 11 : Documentation technique
  - Article 12 : Tenue de registres
  - Article 13 : Transparence et information des utilisateurs
  - Article 14 : **Contrôle humain**



Brian Christian, The alignment problem - How can machines learn human values (Le problème de l'alignement - Comment les machines peuvent-elles apprendre les valeurs humaines), Atlantic Books, 2020

3. L'IA et l'égalité des sexes dans la pratique

- IA, préjugés et égalité des sexes
- Autorités publiques
- Entreprises privées
- Questions juridiques
- Jurisprudence

Égalité des sexes, IA et préjugés

***Préjugés* sexistes**

- **Word2vec/vecteurs de mots :** langage, apprentissage automatique, fréquence, classement, métaphores, importance des ensembles de données + GMLs
- **Stéréotypes et préjugés :** "L'homme est au programmeur informatique ce que la femme est à la ménagère ?" (voir Bolukbasi/Chang et.al. (2016) debiasing word embeddings)
- Les hommes devraient également être associés au congé parental, à l'aide à domicile, au temps partiel...

***Les préjugés* sont humains**

- Les **données** reflètent la société mais façonnent également les perceptions et les préjugés.
- Projet Gender Innovations de Stanford (Machine Learning : Analyse du genre, https://genderedinnovations.stanford.edu)
- Recherche sur le test des associations implicites (Greenwald et Banaij) + implicit.harvard.edu/

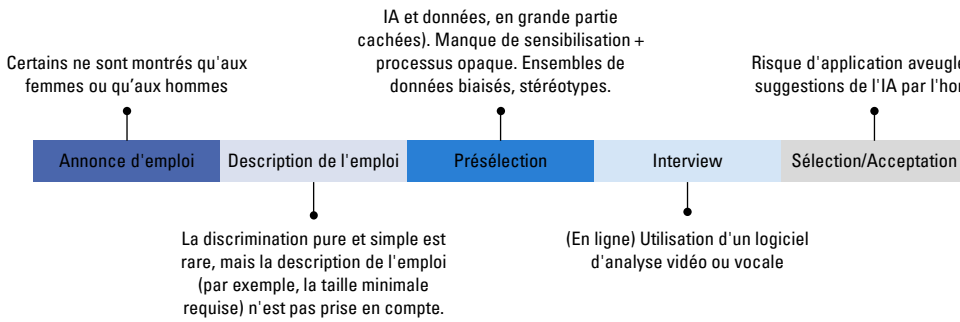
L'IA et l'égalité dans la pratique

Agence pour l'emploi AT

- Utilisation de l'IA par l'État / monopole du placement sur le marché du travail (AMS)
- Discrimination éventuelle à l'encontre des femmes
- Problème : la base de données de l'utilisateur (équilibre vie pro/perso) est reflétée dans les données
- Code source de l'algorithme divulgué

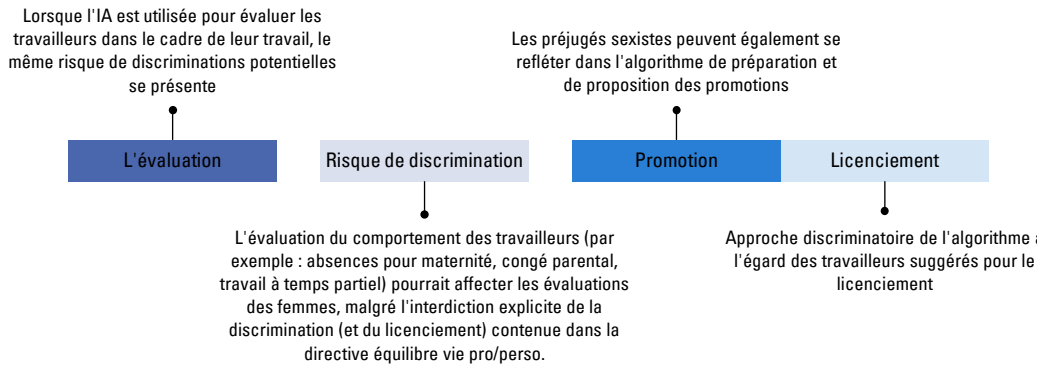
Amazon Recrutement IA

- Acteur privé
- Discrimination éventuelle à l'encontre des femmes
- Problème : données historiques (beaucoup plus de données sur les hommes que sur les femmes)
- Solution : Utiliser de nouvelles données !



L'IA dans le cycle de vie du processus de recrutement (I) - Obtenir l'emploi

25



L'IA dans le cycle de vie du processus de recrutement (II) - Progression de carrière et licenciement

26

Questions juridiques

Problèmes d'évidence et de preuve

- Fournir des preuves (l'IA ne rédige pas de motif de rejet d'un candidat)
- Charge de la preuve (qui prouve quoi ?) + renversement
- Accès à des systèmes d'IA opaques ?
- L'algorithme en tant que secret commercial/protection juridique commerciale
  - Le processus *à l'intérieur de l'appareil photo* comme solution
- Juge en tant qu'expert en IA ou expertise externe nécessaire ?

Application de la loi

- Commission européenne, DG Justice
- Organismes nationaux de lutte contre la discrimination (voir nouveau Dir)
- Application privée
- Procédure de renvoi conformément à l'art. 267 TFUE (importance capitale pour le développement du droit)
  - Quand un tribunal national saisira-t-il la CJCE d'une affaire de discrimination impliquant l'IA ?
- Scientifique en chef des données et de l'informatique pour KOM

Jurisprudence de l'UE (I)

Danfoss C-109/88 (1988)

- "lorsqu'une entreprise **applique** un système de rémunération **manquant totalement de transparence**, il appartient à l'employeur de prouver que sa pratique en matière de salaires n'est pas discriminatoire, si une travailleuse établit, par rapport à un nombre relativement important de salariés, que la rémunération moyenne des femmes est inférieure à celle des hommes" (§ 16)

Kelly, C-104/10 (2011)

- ne permet **pas** à un candidat à la formation professionnelle, qui estime que sa candidature n'a pas été retenue en raison d'une violation du **principe de l'égalité de traitement, d'obtenir des informations détenues par le prestataire de cours sur les qualifications des autres candidats au** cours en question, afin **qu'il puisse établir "des faits qui permettent de présumer l'existence d'une discrimination directe ou indirecte"**, conformément à cette disposition. "(§ 38)
- "Néanmoins, il ne saurait être exclu qu'un refus de divulgation de la part du défendeur, dans le cadre de l'établissement de tels faits, risque de compromettre la réalisation de l'objectif poursuivi par ladite directive et de priver ainsi, notamment, l'article 4, paragraphe 1, de celle-ci de son effet utile. Il appartient à la juridiction de renvoi de vérifier si tel est le cas dans l'affaire au principal. " (§ 39)

Jurisprudence de l'UE (II)

Meister C-415/10 (2012)

- Pas de droit à l'information sur les autres candidats en cas de non-sélection (§ 46)
- Toutefois, il n'est pas exclu que **le refus d'un défendeur d'accorder tout accès à l'information soit l'un des éléments à prendre en compte dans le cadre de l'établissement de faits permettant de présumer l'existence d'une discrimination directe ou indirecte.** Il appartient à la juridiction de renvoi (...) de déterminer si tel est le cas dans l'affaire au principal." (§ 47)

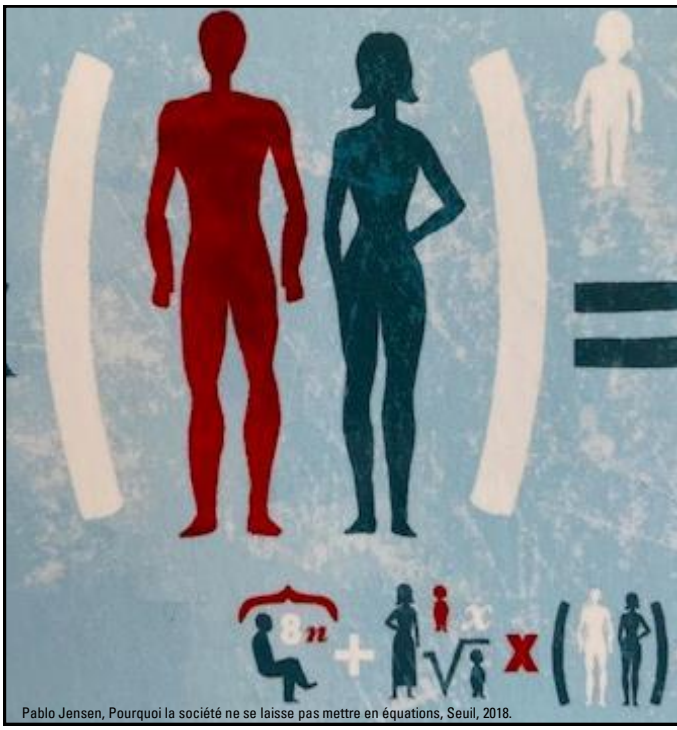
Shuch-Ghannadan, C-274/18 (2019).

- il est **établi que** cette législation affecte négativement un **pourcentage significativement plus élevé de travailleurs féminins que de travailleurs masculins** et si cette législation n'est pas objectivement justifiée par un objectif légitime ou si les moyens de réaliser cet objectif ne sont pas appropriés et nécessaires. L'article 19, paragraphe 1, de cette directive doit être interprété en ce sens qu'il n'impose pas à la partie qui s'estime lésée par une telle discrimination de présenter, en vue d'établir une présomption de discrimination**, des statistiques ou des faits précis relatifs aux travailleurs concernés par la législation nationale en cause si cette partie n'a pas accès à ces statistiques ou à ces faits ou ne peut y accéder que difficilement.**

Jurisprudence de l'UE (III) : Les limites juridiques de l'IA ?

- **ARRÊT DE LA CJEU DU 21. 6. 2022 - AFFAIRE C-817/19 LIGUE DES DROITS HUMAINS ECLI:EU:C:2022:491**
  - Définition de l'IA et de l'apprentissage automatique (paragraphes 58, 194 et 195)
  - Quelles sont les limites ? Pour le contexte PNR (exige des critères "prédéterminés"), la non-stabilité des critères inhérents aux algorithmes d'apprentissage automatique a été décisive (" (...) cette exigence exclut l'utilisation de la technologie de l'intelligence artificielle dans les systèmes d'auto-apprentissage ("apprentissage automatique"), capables de modifier sans intervention ou examen humain le processus d'évaluation (...) ", para. 194)
  - Que fera la Cour dans le contexte des droits fondamentaux ? Des décisions à forts enjeux ?
  - Quelles sont les perspectives pour le principe de non-discrimination ? La CJUE a pris connaissance de l'avis AI/ML + AG du 27/01/22 (§§ 2,228) → *En attendant.... le* prochain arrêt (Art. 267 TFUE)

L'IT interdit l'utilisation de ChatGPT en raison de violations du RGPD et d'autres agences de l'UE étudient la question.
→ Le RGPD est souvent considéré comme une loi AD par les chercheurs et les praticiens (souvent en raison de l'absence de règles spécifiques concernant l'IA et les questions d'égalité des sexes et de discrimination).
→ Interaction entre différents régimes juridiques RGPD (voir Lütz, La pollinisation croisée entre droit de la protection des données et droit de la non-discrimination, *à paraître en 2023*)



Pablo Jensen, Pourquoi la société ne se laisse pas mettre en équations, Seuil, 2018.

4. Réglementation des algorithmes

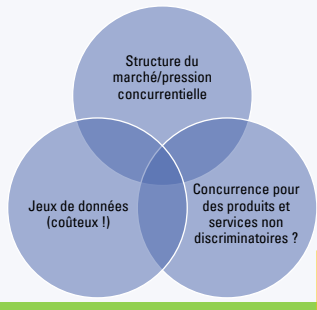
- Approches réglementaires
- Outils réglementaires et détection des préjugés et de la discrimination
- Juridiction de l'UE
- Regarder au-delà des frontières de l'UE

ERA (Trèves 12-13 juin 2023)

31

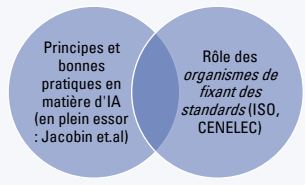
Règlement

Marché



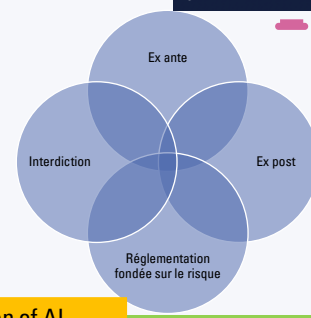
Risque de fixation de règles privées ?

L'autorégulation

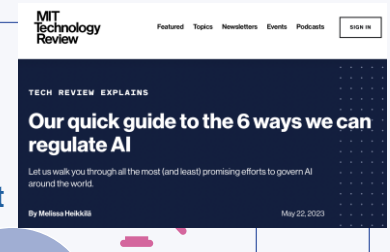


Jurowetski et al, The Privatization of AI Research(-ers) : Causes and Potential Consequences (2021)

Règlement



Auto-évaluations ou audits de l'IA ?



ERA (Trèves 12-13 juin 2023)

32

32

Rôle de la diligence des entreprises en matière de droits de l'homme dans l'IA

Déclaration de Toronto

identifie trois étapes essentielles pour la diligence des entreprises en matière de droits de l'homme pour les systèmes d'apprentissage automatique :

- (a) l'**identification** des résultats discriminatoires potentiels
- (b) la **prévention et l'atténuation de la** discrimination et le suivi des réponses
- (c) la **transparence** des efforts déployés pour identifier, prévenir et atténuer la discrimination

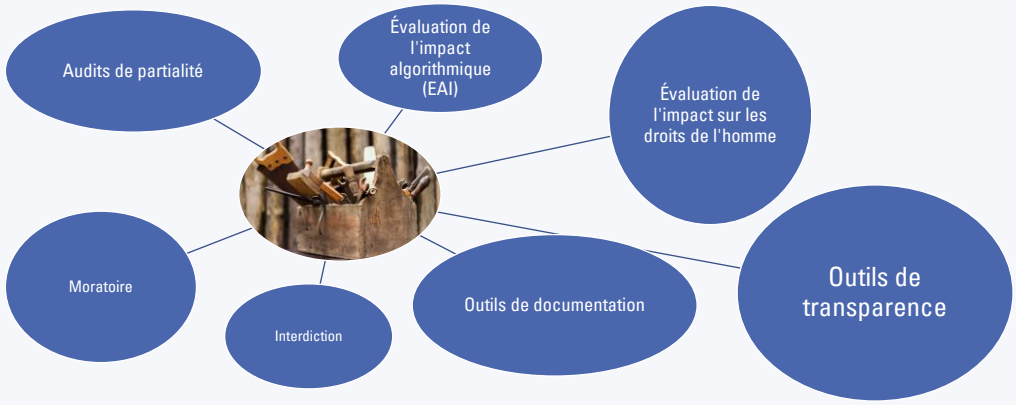
33

Niveau de l'entreprise et rôle des chercheurs

- Les entreprises d'IA devraient utiliser des outils tout au long du cycle de vie de l'IA pour garantir l'équité et/ou détecter les préjugés dans leurs algorithmes/IA.
  - Guides spécifiques pour les développeurs et les programmeurs d'IA
  - Éducation et formation pour garantir des algorithmes justes et impartiaux
- **Les chercheurs** (indépendants) sont essentiels pour évaluer les besoins concrets et la faisabilité des obligations réglementaires , par exemple pour détecter, réduire et éliminer les biais de la conception et des ensembles de données afin de réduire les résultats discriminatoires.

34

Outils permettant d'atteindre les objectifs réglementaires : atténuation des préjugés et réduction de la discrimination



35

Dimension mondiale

Le Conseil de l'Europe

- Travaux en cours pour l'élaboration d'un cadre législatif (2025)
- Recommandation CM/Rec(2020)1 sur l'impact des systèmes algorithmiques sur les droits de l'homme
- "Nécessité de veiller à ce que (...) les **déséquilibres entre les sexes et les autres déséquilibres au sein de la main-d'œuvre qui n'ont pas encore été éliminés de nos sociétés ne soient pas délibérément ou accidentellement perpétués par des systèmes algorithmiques**" (Préambule)

OCDE

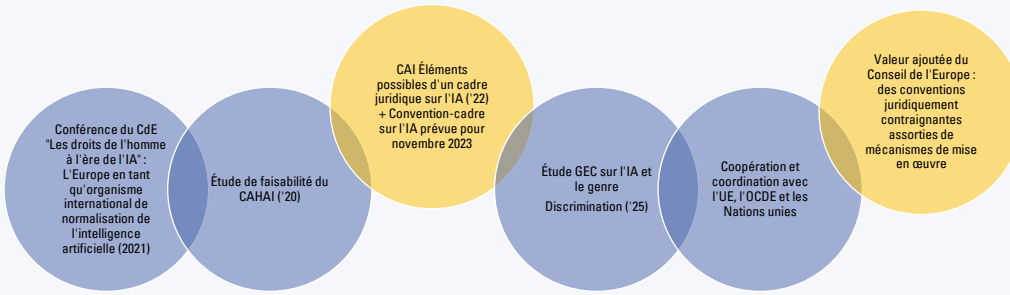
- OECD/LEGAL/0449, Recommandation du Conseil sur l'intelligence artificielle (2019)
- "Les acteurs de l'IA devraient respecter (...) les droits de l'homme et les valeurs démocratiques, tout au long du cycle de vie du système d'IA. Il s'agit notamment (...) de la **non-discrimination et de l'égalité** (...)" (1.2a)

Nations Unies

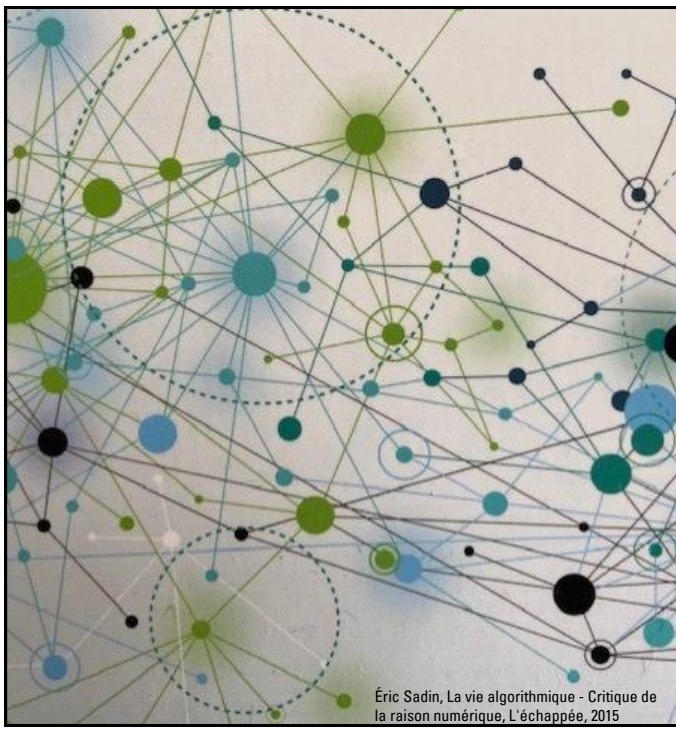
- Recommandation de l'UNESCO sur l'éthique de l'IA (2021), Genre comme domaine d'action 6 (§§ 87-93)
- Le droit à la vie privée à l'ère numérique, Rapport du Haut Commissaire des Nations unies aux droits de l'homme, A/HRC/48/31 :
- "Les progrès des nouvelles technologies ne doivent pas être utilisés pour **éroder les droits de l'homme, creuser les inégalités ou exacerber les discriminations existantes**" (§ 4) + "**Ces systèmes (...) décident qui a une chance d'être recruté pour un emploi**" (Rapport, § 57).

36

Conseil de l'Europe



37



Éric Sadin, La vie algorithmique - Critique de la raison numérique, L'échappée, 2015

5. L'IA, catalyseur de l'égalité entre les hommes et les femmes ?

- L'IA, détective de la discrimination
- Comparaison humain / IA
- Considérations pratiques

38

L'IA, détective de la discrimination

Éléments

- Du visible à l'invisible, de l'explicite à l'opaque
- Les décisions humaines sont sujettes à erreur, mais même l'IA n'est pas nécessairement meilleure
- Les algorithmes peuvent surmonter (certains) *biais* et le *bruit* (K/S/S, Bruit, p. 334-337).

Exemples

- Analyse automatique de la jurisprudence
- Une application mieux ciblée, une meilleure utilisation des ressources
- Utiliser des algorithmes pour tester d'autres algorithmes en vue d'éventuelles discriminations
- Logiciel de détection des biais et de la discrimination

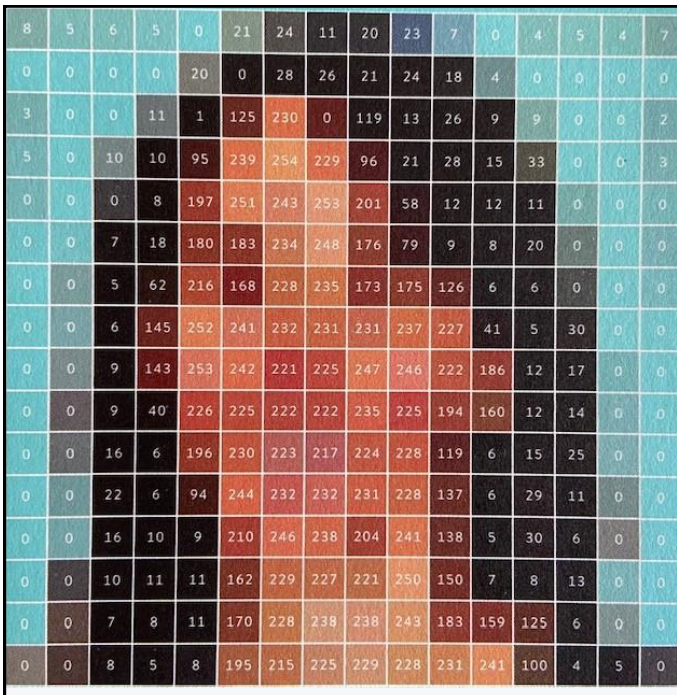
Comparaison entre l'homme et l'IA

Avantages de l'IA

- Cohérence, objectivité, rapidité ?
- De meilleurs décideurs ?
- Exercice du pouvoir discrétionnaire ?
- Identification des modèles
- L'IA est d'une (grande) aide (soutien) mais les humains doivent garder le contrôle

L'IA comme assistant

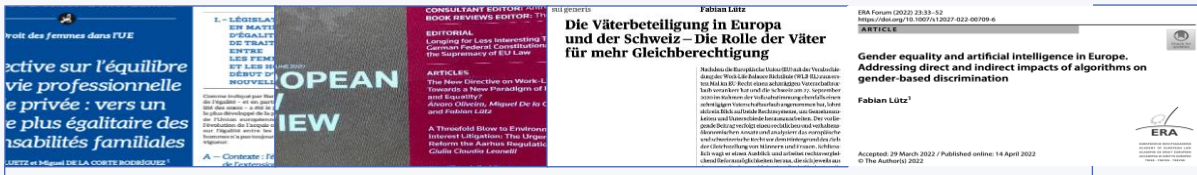
- Assistant du "bureaucrate de base" (Lipsky, 1980)
- Plus de prévisibilité et de fiabilité dans les décisions administratives ?
- Préparation de l'analyse de cas
- Conclusion : L'IA ne peut pas remplacer la prise de décision
- *"Le pouvoir sans précédent de l'intelligence artificielle qui peut être une **force pour le bien** (...) (Handicap (A/HRC/49/52))*



Melanie Mitchell, Artificial intelligence - A guide for thinking humans, Pelican, 2019.

6. Conclusion et perspectives

- L'IA présente des risques pour l'égalité entre les hommes et les femmes
- Opportunités également
- La réglementation est essentielle
- Rôle partagé des entreprises et des États
- Dimension mondiale



Publications choisies sur les algorithmes et l'égalité des sexes :

- ***How the 'Brussels Effect' Could Shape the Future Regulation of Algorithmic Discrimination***, Duodecim Astra, Issue 1, p. 142-163 (2021), (Disponible en ligne).
- ***Discrimination by correlation - Towards eliminating algorithmic biases and achieving Gender Equality***, contribution à une monographie, Transcript Verlag (2021), (Open Access).
- ***L'égalité des sexes et l'intelligence artificielle en Europe. Aborder les impacts directs et indirects des algorithmes sur la discrimination fondée sur le genre***. Forum ERA (2022), disponible à l'adresse : https://doi.org/10.1007/s12027-022-00709-6 (Open Access)

- ***Artificial Intelligence and Gender-Based Discrimination***, in : Quintavilla Alberto, Temperman Jeroen (eds.) Human Rights and Artificial Intelligence, Oxford University Press (juin 2023).
- ***Algorithmische Entscheidungsfindung aus der Gleichstellungsperspektive - Ein Balanceakt zwischen Gender Data Gap, Gender Bias, Machine Bias und Regulierung***, GENDER 1/2 (2023), (Open Access online).
- ***L'égalité des sexes et l'intelligence artificielle : SDG 5 and the role of the UN in fighting stereotypes, biases and gender discrimination***, in : Cristani Federica, Fornalé Elisa (eds.) Women's Empowerment and its Limits, Palgrave, mai 2023).
- ***Le rôle du droit pour contrer la discrimination algorithmique dans le recrutement automatisé***, In : Guillaume Florence (eds.) La technologie, l'humain et le droit, Stämpfli Verlag (juin 2023)

REMERCIEMENTS
à l'ERA

MERCI
de m'avoir écouté

DANKE
pour l'intérêt porté à la discrimination algorithmique

Fabian.Luetz@unil.ch