Réglementer l'IA ?
Perspectives pour la Suisse

académies suisses
des sciences

SCIENCE ET POLITIQUE

à table !

Programme

- **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
- **Que signifie l'AI Act de l'UE pour la Suisse ?**
  Nadja Braun Binder, professeure de droit public à l'Université de Bâle
- **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
- **Discussion**

UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG

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

Prof. Dr. Philippe Cudré-Mauroux

eXascale Infolab
Université de Fribourg

eXascale Infolab

Science et politique à table
03.12.2024

Qu’est-ce que l’IA moderne?

Intelligence Artificielle (IA)

ensemble de théories et de techniques visant à réaliser des machines capables de simuler l'intelligence humaine

Apprentissage automatique

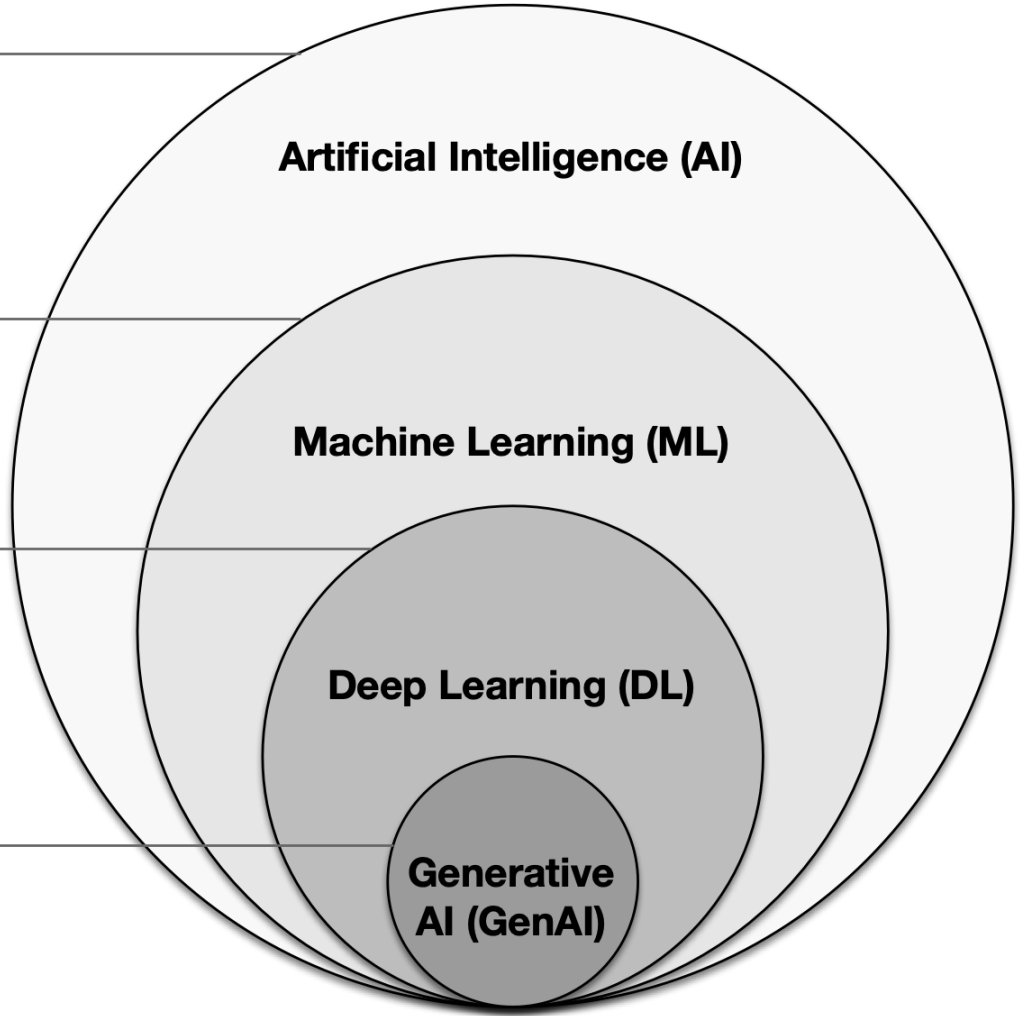
approches statistiques entraînant des modèles à partir de données

Apprentissage profond

apprentissage automatique utilisant des réseaux de neurones artificiels profonds pour apprendre des modèles complexes à partir de mégadonnées

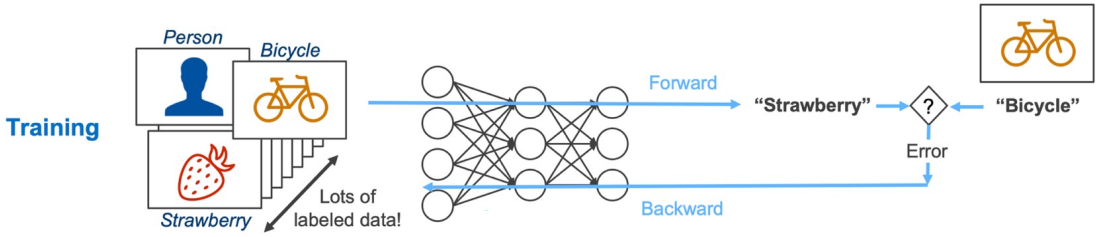
IA générative

sous-ensemble de l’apprentissage profond dont le but est de générer de nouvelles données

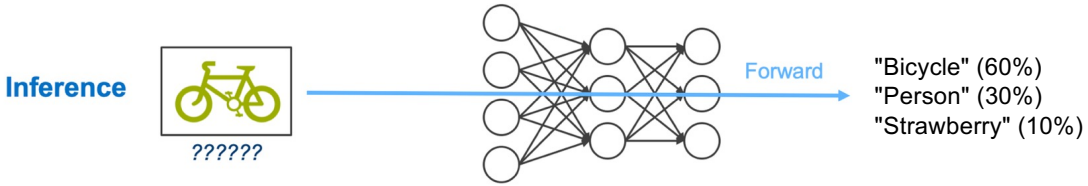


Training & Inference

Les réseaux de neurones artificiels sont dans un premier temps **entraînés** sur des millions d'exemples.



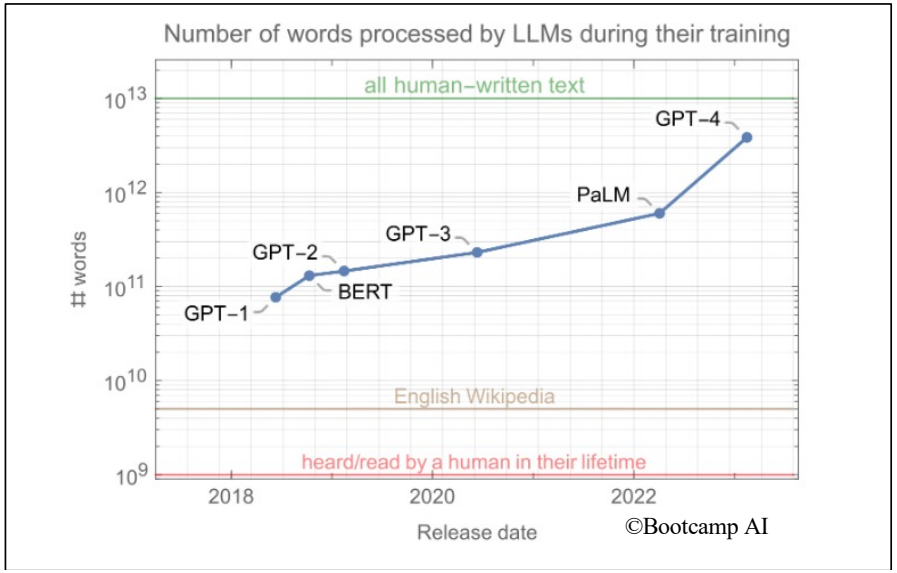
Une fois entraînés, ils sont utilisés pour faire des prédictions (**inférences**) sur de nouvelles données.



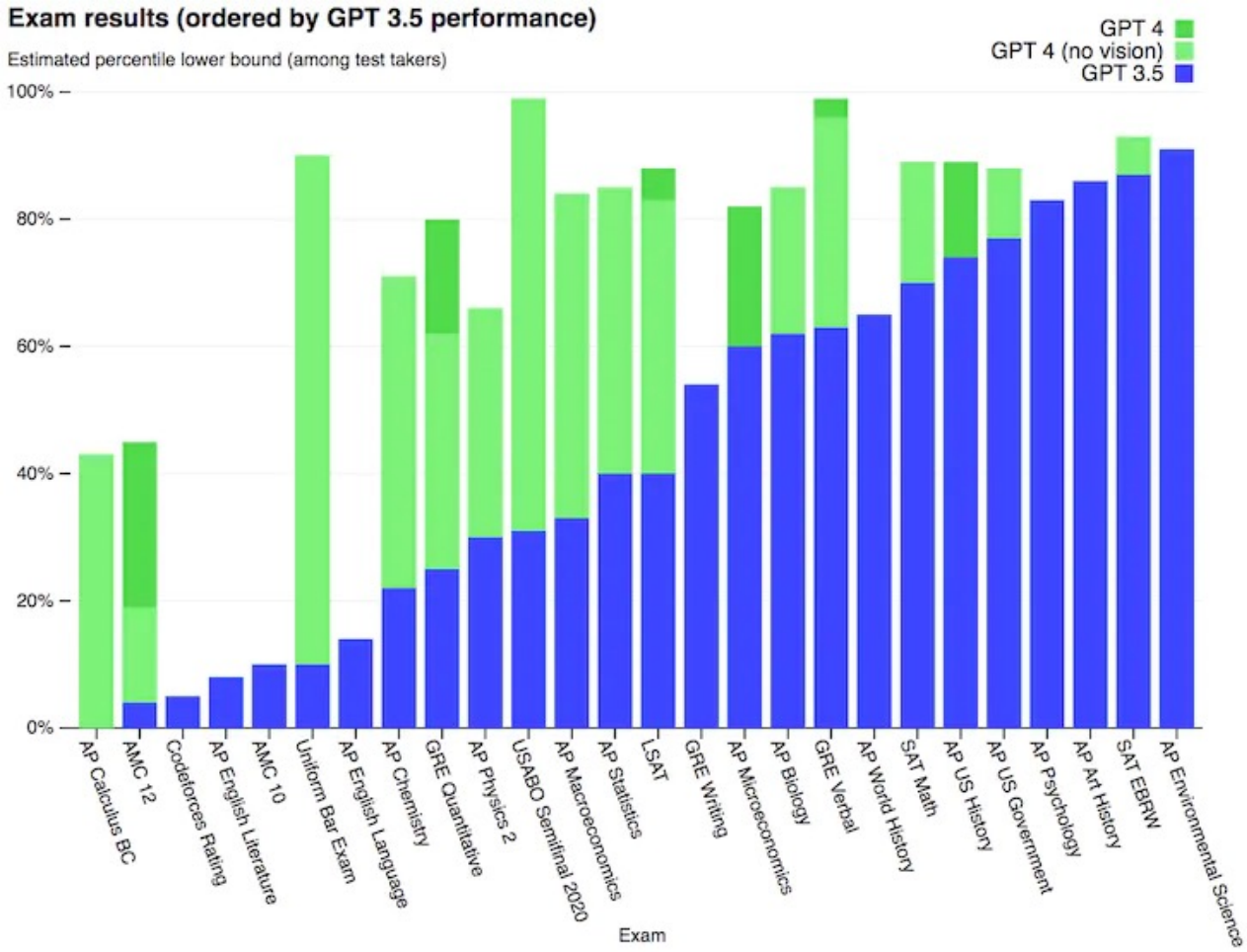
Foundation & Large Language Models

Les modèles récents les plus performants (comme GPT-4)

- ont des centaines de millions de **paramètres**
- entraînés sur des jeux de **données** colossaux
- en utilisant des dizaines de milliers de **GPUs**
- pour des coûts se chiffrant en dizaines voire centaines de **millions de dollars.**



… amenant des résultats spectaculaires



… et des résultat malheureusement peu fiables

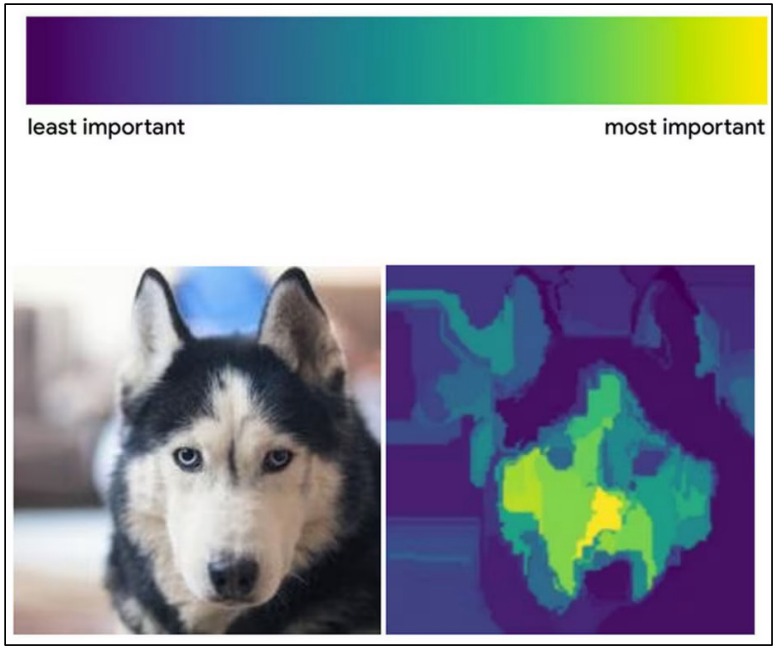
- **incertitude**: les modèles retournent toujours plusieurs résultats probables
- **overfitting**: les modèles produisent des résultats hasardeux pour des cas nouveaux (non vus durant l'entraînement)
- **biais**: les modèles reflètent les biais de leurs données
- **hallucinations**: les modèles de GenAI génèrent des contenus nouveaux, qui ne sont pas forcément des faits

Une IA fiable – possible ou non?

- Il est difficile (voire impossible) techniquement d'éliminer ces problèmes
- Par contre, la recherche académique développe depuis plusieurs années des techniques prometteuses pour les **détecter**, les **amenuiser** ou les **contrebalancer**

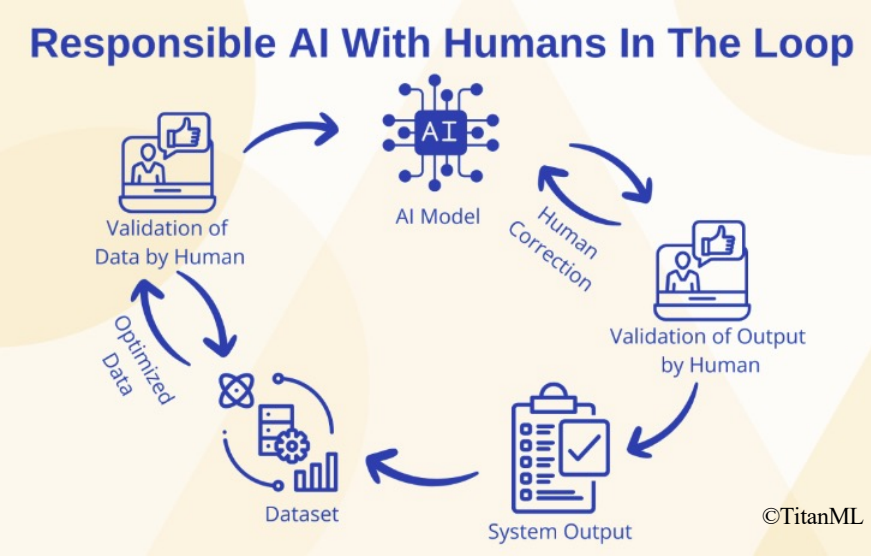
eXplainable AI (xAI)

- Les techniques de l'IA explicable sont utilisées pour décrire un modèle d'IA, son impact attendu et ses biais potentiels.



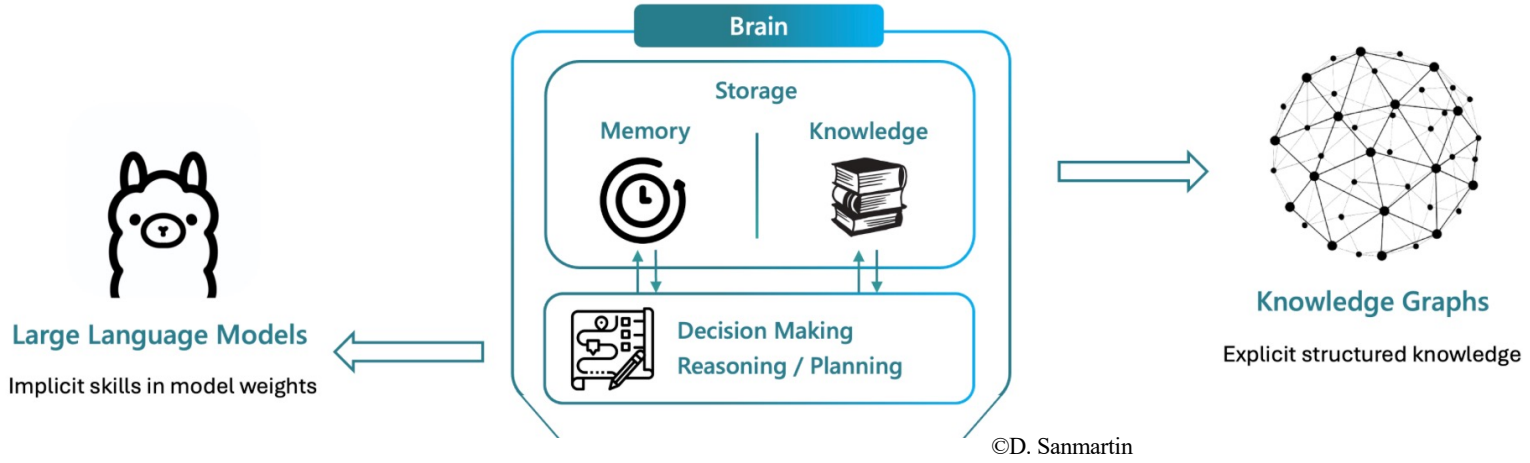
Human-in-the-Loop AI

- Intègre de manière systématique et continue du feedback humain afin d'améliorer la précision, la fiabilité, et l'adaptabilité des modèles



Neuro-Symbolic AI

- Combine des architectures d'IA neuronales et symboliques (utilisant par ex. des règles logiques ou des graphes de connaissances), pour fournir une IA plus robuste



Conclusion

- Les modèles d'IA moderne sont **extrêmement larges** et **complexes**, très **performants** mais **peu fiables**
- Ce manque de fiabilité est **intrinsèque** aux méthodes d'entraînement utilisées et ne va pas disparaître
- Par contre, la recherche académique développe des techniques prometteuses pour **détecter**, **amenuiser** ou **contrebalancer** ces problèmes

Merci pour votre attention!

https://exascale.info

Universität Basel

Qu'implique la loi de l'UE sur l'IA pour la Suisse ?

Prof. Dr Nadja Braun Binder

03.12.2024

ePIAF
Electronic Public Institutions and Administrations Research Forum

1. Situation initiale et conditions-cadres

Aktuelles
Europäisches Parlament

Ursula von der Leyen stellt dem Plenum ihre Leitlinien vor

Pressemitteilung PLENARTAGUNG 16-07-2019 - 14:26

A Europe fit for the digital age

«**In my first 100 days in office, I will put forward legislation for a coordinated European approach on the human and ethical implications of Artificial Intelligence**. This should also look at how we can use big data for innovations that create wealth for our societies and our businesses.»

https://www.europarl.europa.eu/news/fr/press-room/20190711IPR56823/ursula-von-der-leyen-a-presente-son-programme-aux-deputes

1. Situation initiale et conditions-cadres

Eléments de la prise en compte dans le droit de l'IA dans l'UE:

- **L'accès à des données de grande qualité** est le principal facteur de systèmes d'IA performants et robustes
  - Loi sur les données
  - Règlement sur la gouvernance des données
- **Confiance en l'IA**
  - Cadre juridique européen pour l'IA
  - Cadre de responsabilité civile
  - Révision de la législation sectorielle en matière de sécurité
- Cadre juridique **existant**
  - p.ex. Règlement européen sur la protection des données (RGPD)

1. Situation initiale et conditions-cadres

Objectifs du règlement sur l'IA (→ art. 1 du règlement sur l'IA)

- Amélioration du fonctionnement du **marché intérieur** européen
- Promotion de l'IA **axée sur l'humain** et **digne de confiance**
- Niveau élevé de protection en ce qui concerne la **santé**, la **sécurité** et **les droits fondamentaux** – y compris la démocratie, l'état de droit et la protection de l'environnement – contre les effets néfastes
  (→ sécurité des produits)

2. Dispositions principales du règlement sur l'IA

Définition du système d'IA

Art. 3 Définitions

« Aux fins du présent règlement, on entend par

1. « système d'IA », un système automatisé qui est conçu pour fonctionner à **différents niveaux d'autonomie** et peut faire preuve d'une **capacité d'adaptation** après son déploiement, et qui, pour des objectifs explicites ou implicites, déduit, à partir des entrées qu'il reçoit, la manière de générer des sorties telles que des **prédictions, du contenu, des recommandations ou des décisions** qui peuvent influencer les environnements physiques ou virtuels ; »

2. Dispositions principales du règlement sur l'IA

Champ d'application personnel du règlement sur l'IA (→ art. 2)

- Fournisseurs, déployeurs, importateurs, distributeurs, fabricants de produits (qui mettent sur le marché ou mettent en service un système d'IA en même temps que leur produit et sous leur propre nom ou leur propre marque)
  - qui ont leur lieu d'établissement ou sont situés dans l'Union
  - Fournisseurs et déployeurs qui ont leur lieu d'établissement ou sont situés dans un **pays tiers**, **lorsque les sorties produites par le système d'IA sont utilisées dans l'Union**
- Personnes concernées qui sont situées dans l'Union

2. Dispositions principales du règlement sur l'IA

Catégorisation des systèmes d'IA

- IA à usage spécifique (single-purpose AI)

→ classification en fonction du risque lié à son **utilisation** :

  - Pratiques **interdites** (art. 5)
  - **Systèmes d'IA à haut risque** (art. 6 ss)
  - Systèmes d'IA à risque **limité** : obligations de transparence (art. 50)
  - Systèmes d'IA à risque **minimal** : pas de prescriptions

2. Dispositions principales du règlement sur l'IA

Catégorisation des systèmes d'IA

- IA à usage général (general purpose AI, GPAI) et modèles de base

→ Classification en fonction de la **performance et de la portée** du modèle de base :

  - GPAI : obligations de transparence supplémentaires
  - IA à usage général **présentant un risque systémique** (art. 51) : obligations de transparence supplémentaires et autres obligations (surveillance, évaluation des modèles, essais contradictoires)

3. Exemple tiré d'une disposition du règlement sur l'IA

Art. 10 du règlement sur l'IA : Données et gouvernance des données

1 (...)

2 Les jeux de données d'entraînement, de validation et de test sont soumis à des pratiques en matière de gouvernance et de gestion des données **appropriées** à la destination du système d'IA à haut risque. (…)

3 Les jeux de données d'entraînement, de validation et de test sont **pertinents**, suffisamment **représentatifs** et, dans toute la mesure possible, **exempts d'erreurs** et **complets** au regard de la destination. (…)

4 Les jeux de données tiennent compte, dans la mesure requise par la destination, des caractéristiques ou éléments **propres** au cadre géographique, contextuel, comportemental ou fonctionnel spécifique dans lequel le système d'IA à haut risque est destiné à être utilisé.

(…)

4. CH : Reprise autonome de la loi sur l'IA ?

- L'UE a une autre priorité **(harmonisation du marché intérieur)** que la Suisse
- Les normes de la loi sur l'IA sont très **vagues** – une concrétisation serait encore nécessaire
- La reconnaissance des **procédures d'évaluation de la conformité** est loin d'être assurée
- Risque de **doubles emplois** avec des règlements sectoriels déjà existants pour la sécurité des produits
- Concrétisation en normes techniques qui seront **de facto** également pertinentes en Suisse
- La réglementation renforce la position sur le marché des « **Big Tech** »

Universität Basel

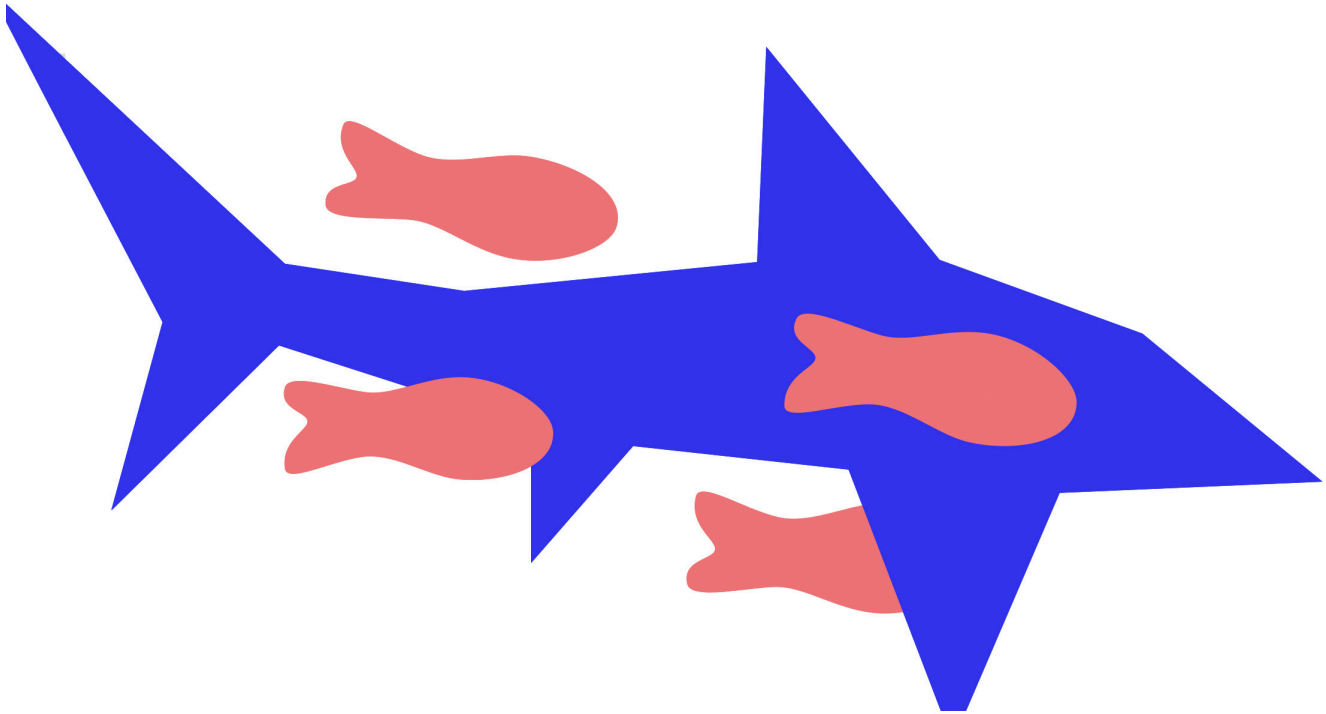
Merci
de votre attention !

ePIAF

Electronic Public Institutions and Administrations Research Forum

Saisir les opportunités de l’IA en gardant le contrôle – comment y parvenir ?

Prof. Dr Thomas Burri, Université de Saint-Gall

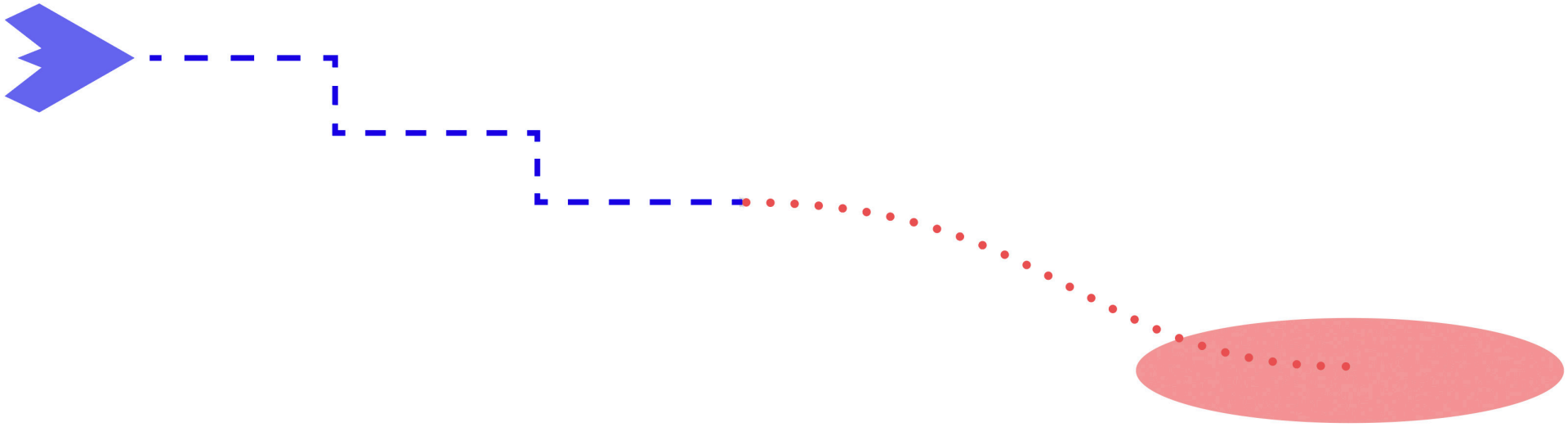


Thomas Burri

Contrôle humain et surveillance humaine

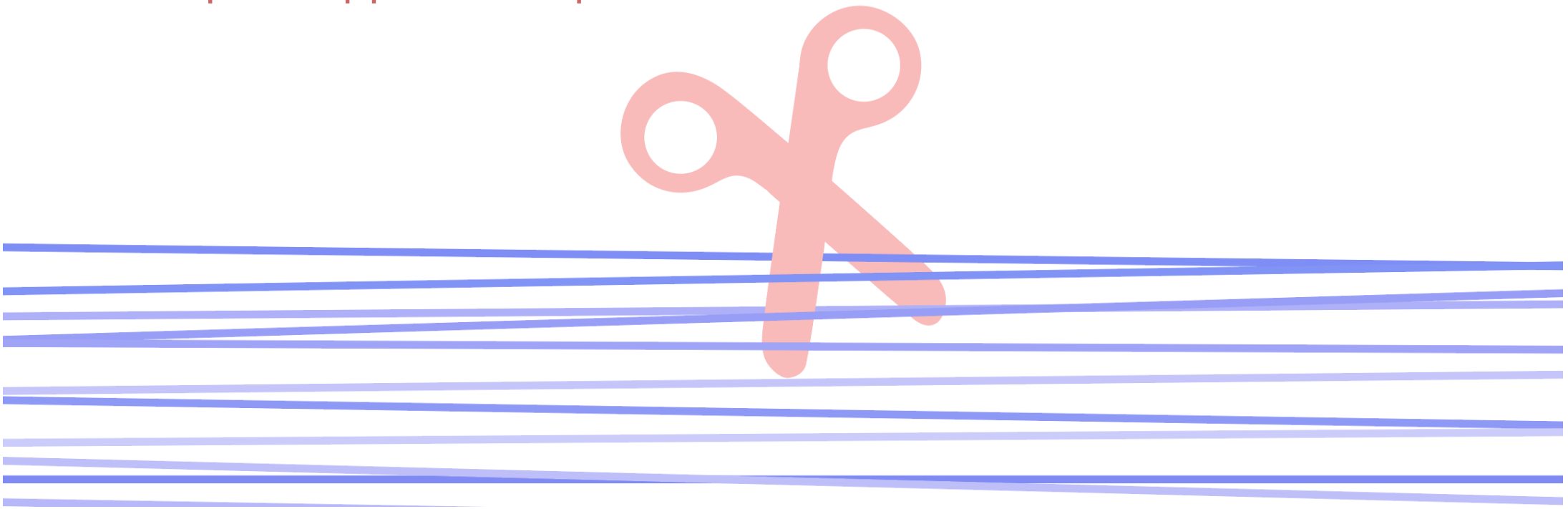
Intelligence artificielle chaude et froide

Atterrissage à l’aide de l’intelligence artificielle



8 Pistes de réflexion

1. Concrétiser dans la pratique les obligations de surveillance de l’IA
2. L’intervention humaine dans l’IA chaude n’est pas la panacée
3. Ne pas supposer l’explicabilité de l’IA moderne



Thomas Burri

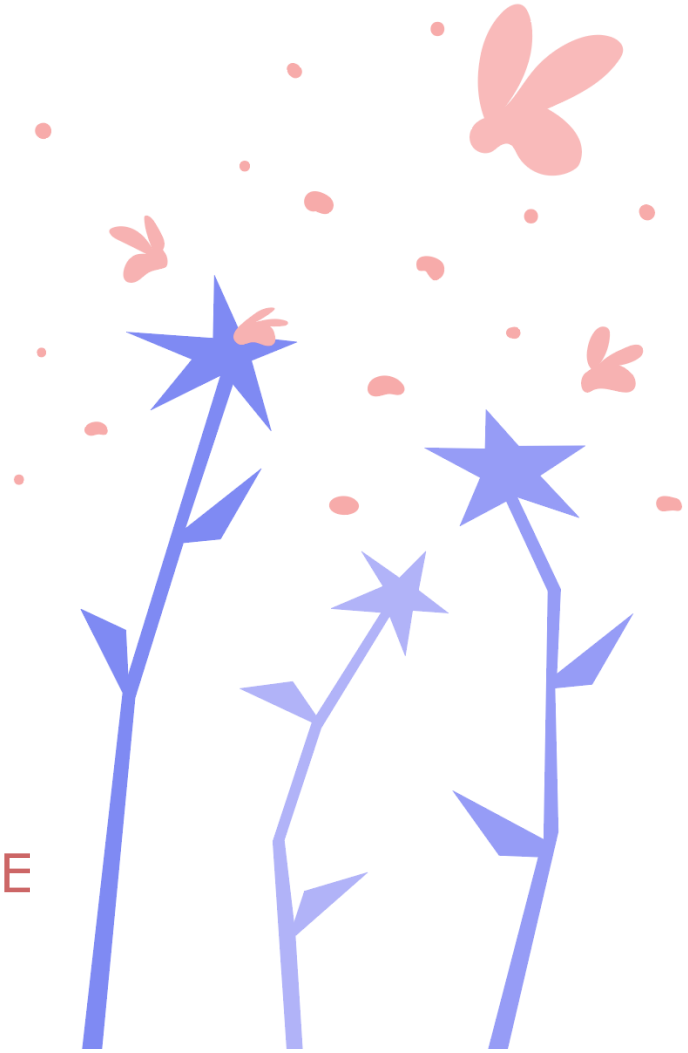
4. La règlementation globale de l’IA a de fortes répercussions sur l’innovation

5. Le droit existant s’applique à l’IA

6. Aborder les répercussions sociales de l’IA séparément des risques opérationnels

7. Approche la plus prometteuse : combiner les principes fondamentaux avec des adaptations spécifiques de la loi

8. Ne pas reprendre le règlement sur l’IA de l’UE pour le moment



Merci !

Prof. Dr Thomas Burri, Université de Saint-Gall

Projet FNS N° 407740_187494
Contrats ArmaSuisse S+T 8003540020, 8003538711, 8003535283

Autres spécialistes (en plus des orateurs)

Alice Delorme Benites, l'Institut de traduction et d'interprétation à la ZHAW

Roger Abächerli, Département des sciences de la santé et de la technologie de l'EPF Zurich; SATW

Emanuela Keller, Professeure et médecin-chef à l'Hôpital universitaire de Zurich

Markus Christen, Digital Society Initiative de l'Université de Zurich, TA-Swiss

académies suisses des sciences

Réglementer l'IA ?
Perspectives pour la Suisse

académies suisses
des sciences

SCIENCE ET POLITIQUE

à table !