

# Et si on démystifiait un peu l'IA générationnelle ?

Bernard Fallery

Professeur émérite de l'université de Montpellier  
Labo Montpellier recherche Management.

Séminaire - Comment l'IA générationnelle transforme nos métiers ?

18 octobre 2024

## Et si on démystifiait un peu l'IA générative ?

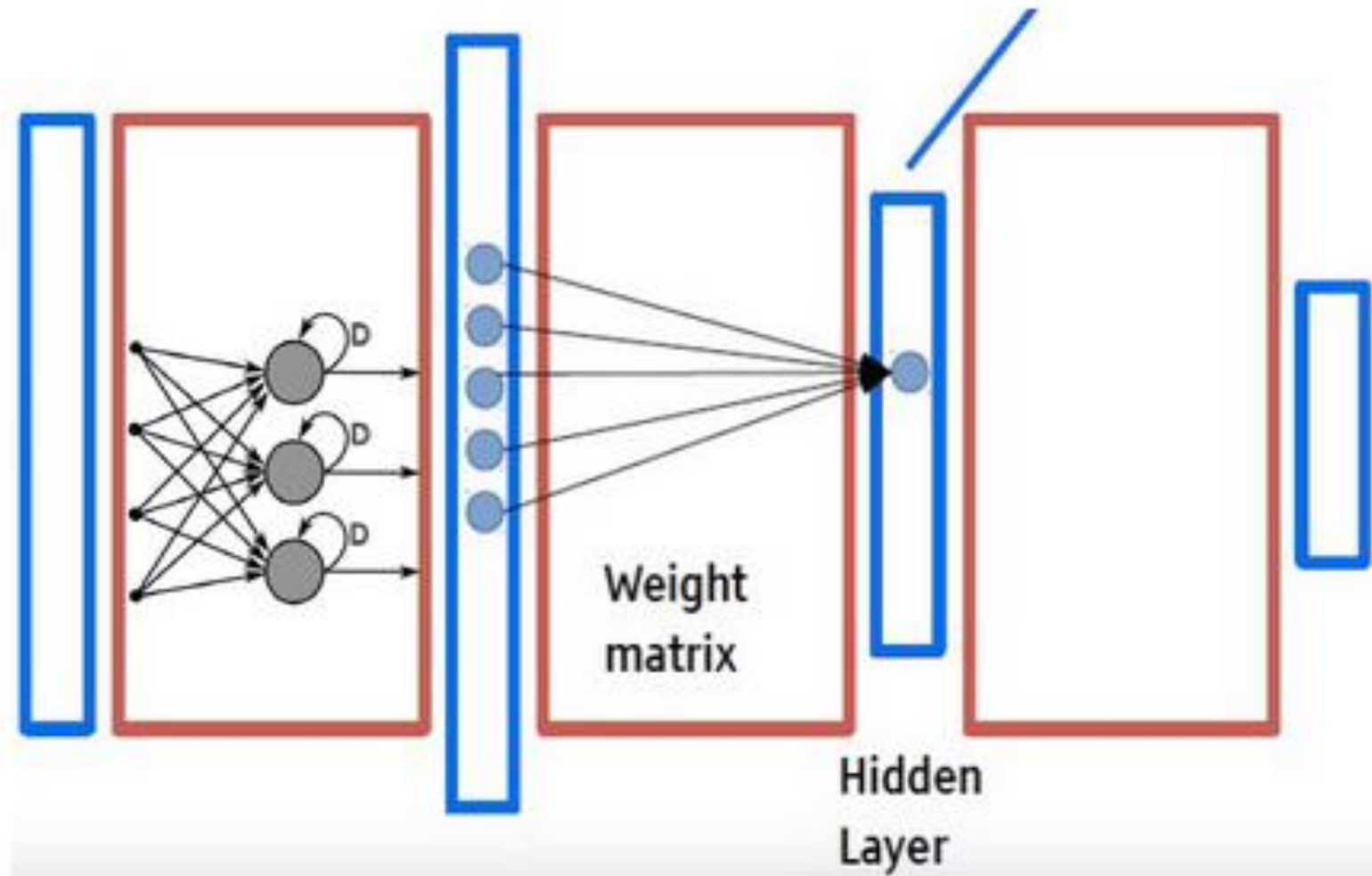
**Questions d'éthique ?** Geoffrey Hinton, Yosuah Bengio... (Pékin 2024) « *La lutte contre les risques liés à l'IA devrait être une priorité mondiale au même titre que d'autres **risques à l'échelle de la société**, tels que les pandémies et les guerres nucléaires »*

**Questions d'environnement ?** Sundar Pichai, Google, 2024  
« *La demande de calcul informatique liée à l'IA a été **multipliée par un million** en six ans »*

**Questions de souveraineté ?** Sam Altman, OpenAI, 2024  
« *OpenAI **pourrait cesser d'opérer** dans l'Union européenne si le futur règlement européen lui imposait trop de contraintes »*

# 1. Affirmer trois phrases non dangereuses

- 1. **L'IA** est une branche de l'informatique qui décompose les fonctions cognitives pour pouvoir **SIMULER** des comportements humains : **symbolique // connexionniste**
- 2. **Un système d'IA** est un logiciel qui, pour des objectifs définis par l'homme, génère des prédictions, des recommandations, des décisions, des contenus... en **INTERAGISSANT avec son environnement** (boucles d'interactions autonomes : dans l'**automatisation** // dans la **vectorisation**)
- 3. **Un système d'IA générative**, issu de Larges Modèles de Langage, est un profond réseau de matrices pré-entraîné qui génère en sortie des **textes ou documents très plausibles**



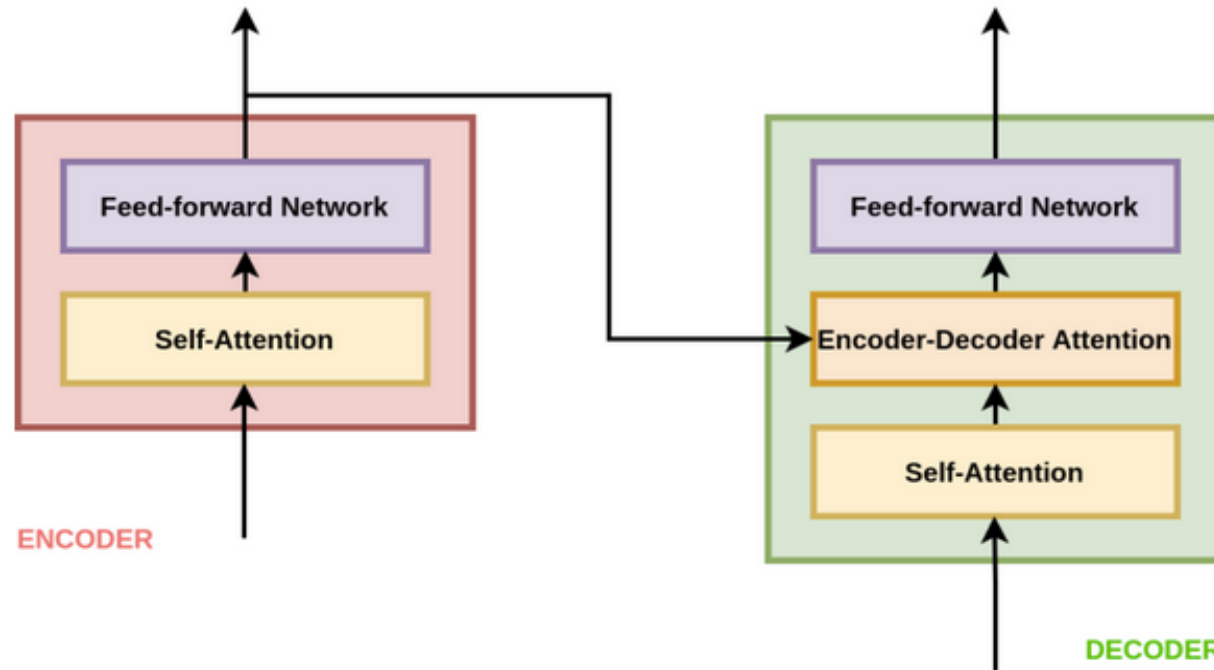
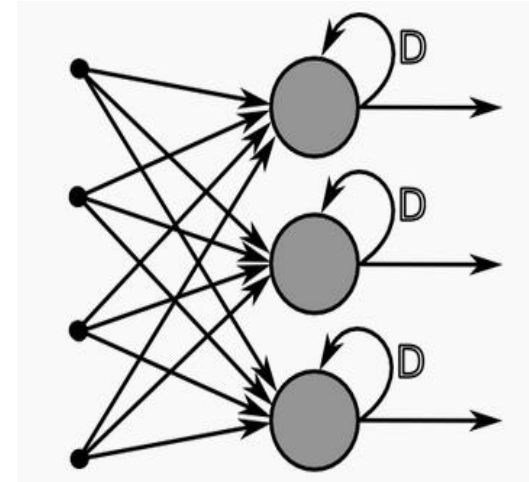
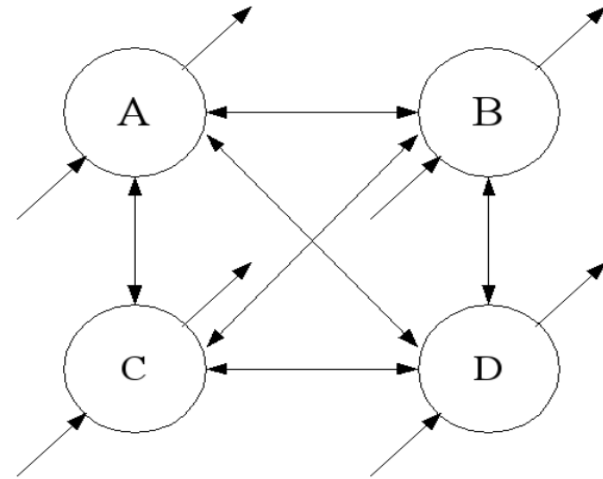
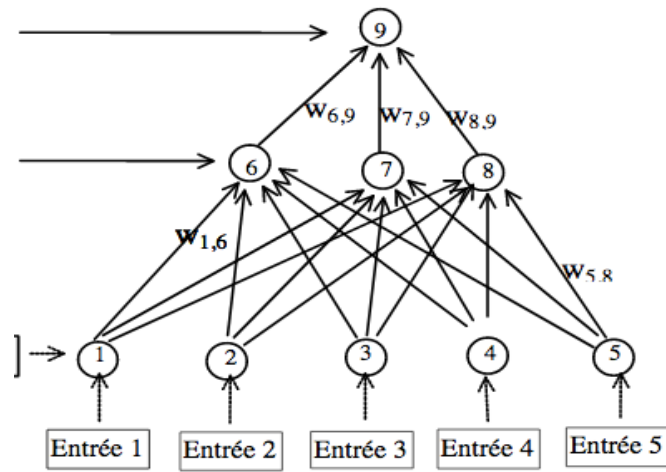
## La retro-propagation du gradient d'erreur

$$E = \sum_{i=1}^s \sum_{j=1}^n \frac{(y_{i,j} - d_{i,j})^2}{2} \quad \Delta w_{i,j} = - \frac{\partial E}{\partial w_{i,j}} \varepsilon$$

## $\alpha$ , self-attention dans les séquences d'entrées

d'attention  $\alpha_{i,j}$  mesurant une similarité entre toute paire  $(i, j)$  de la séquence  $\{\mathbf{x}_t\}_{t=1\dots T}$  (voir également la Figure 2c) :

$$\alpha_{i,j} = \frac{\exp(e_{i,j})}{\sum_{i'=1}^H \exp(e_{i',j})}, \quad e_{i,j} = \frac{(\mathbf{x}_j^T W_Q)^T (\mathbf{x}_i^T W_K)}{\sqrt{D}}$$



Séminaire - Comment l'IA générative transforme nos métiers ?

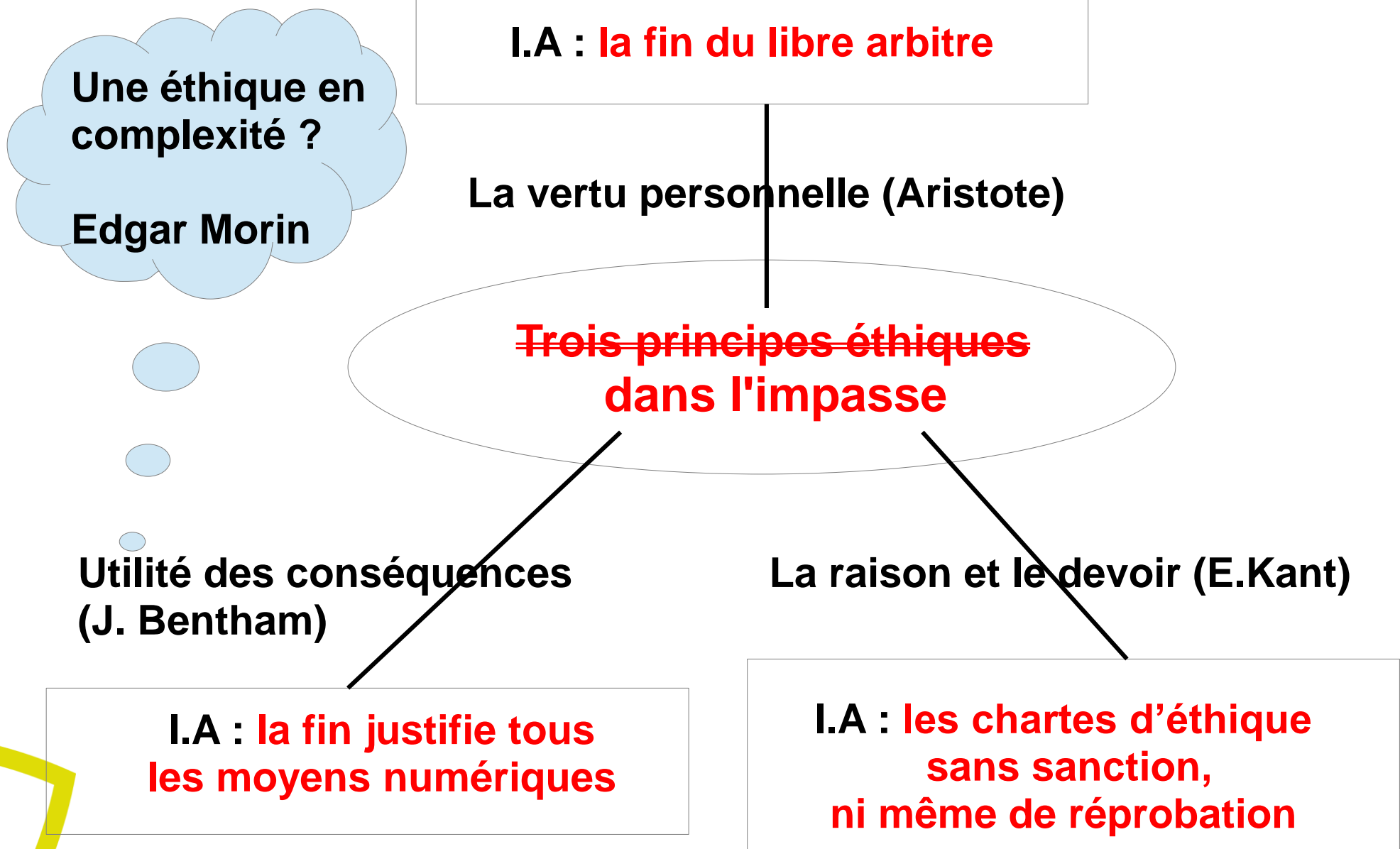
18 octobre 2024

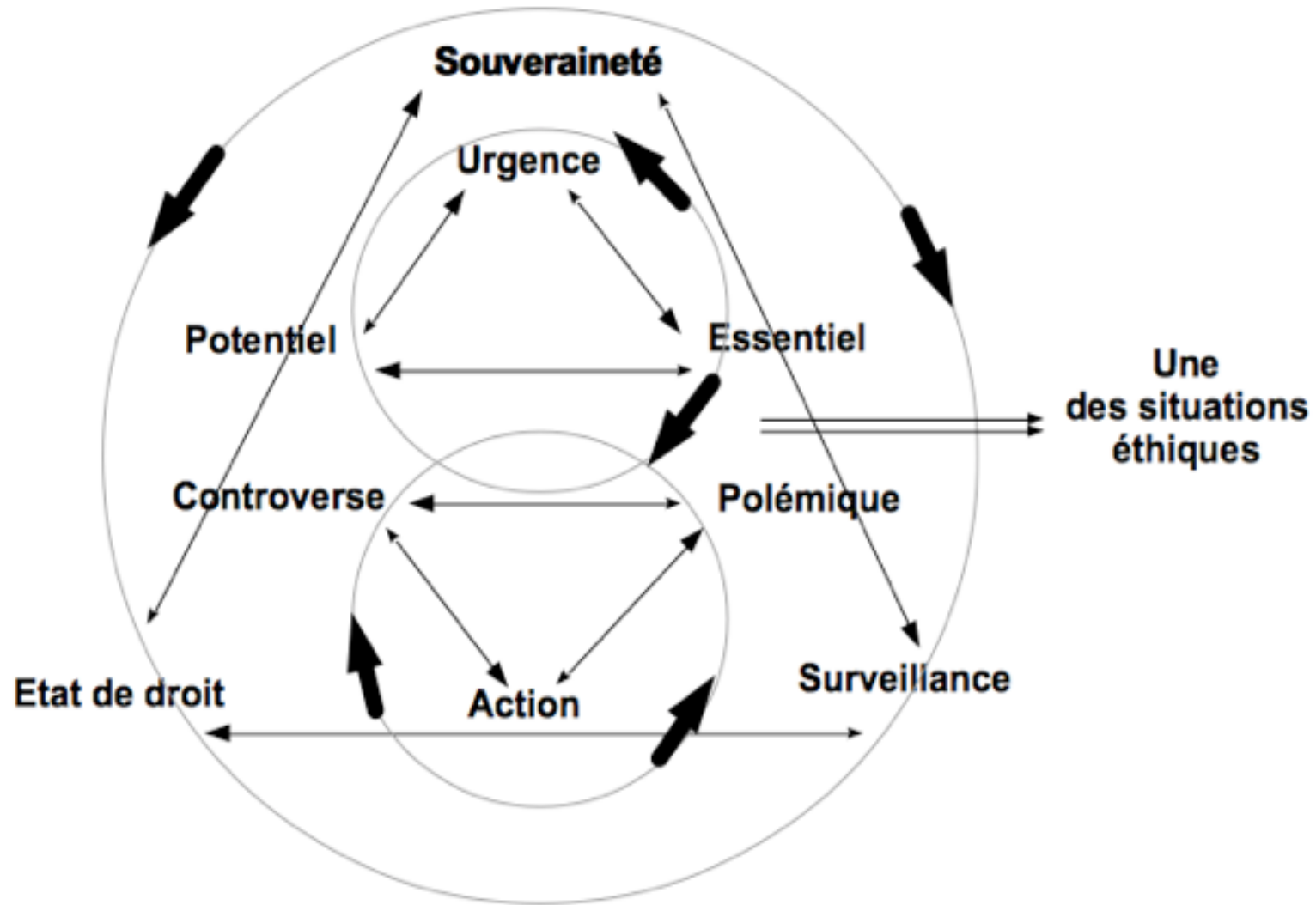
## 2. Contester trois phrases très dangereuses

- a) **L'IA générative n'est qu'un outil**, tout dépend de l'usage qu'on en fait.
- (Les mythes sur le **contrôle** et sur la **productivité**)
- b) Un jour **une IA forte** dépassera l'intelligence humaine.
- (Les mythes sur l'**apprentissage** et sur la **créativité**)
- c) De toute façon, **on n'arrêtera pas l'innovation.**
- (Les mythes sur le **marché** et sur le **solutionnisme**)

- a) « L'IA générative **n'est qu'un outil**, tout dépend de l'usage qu'on en fait »
- Une première confusion sur **le concept d'outil**  
~~Neutralité~~ Outil < Dispositif < Milieu technique
- Le stade du **capitalisme de surveillance**
  - Traces < Profilage < Gouvernamentalité algorithmique.
- Le stade du **techno-féodalisme**
  - Dépendance < Servitude < Économie de la rente
- Une deuxième confusion par **absence d'une réflexion éthique**
- Fin du libre arbitre
- Fin de la déontologie
- Fin de la Justice







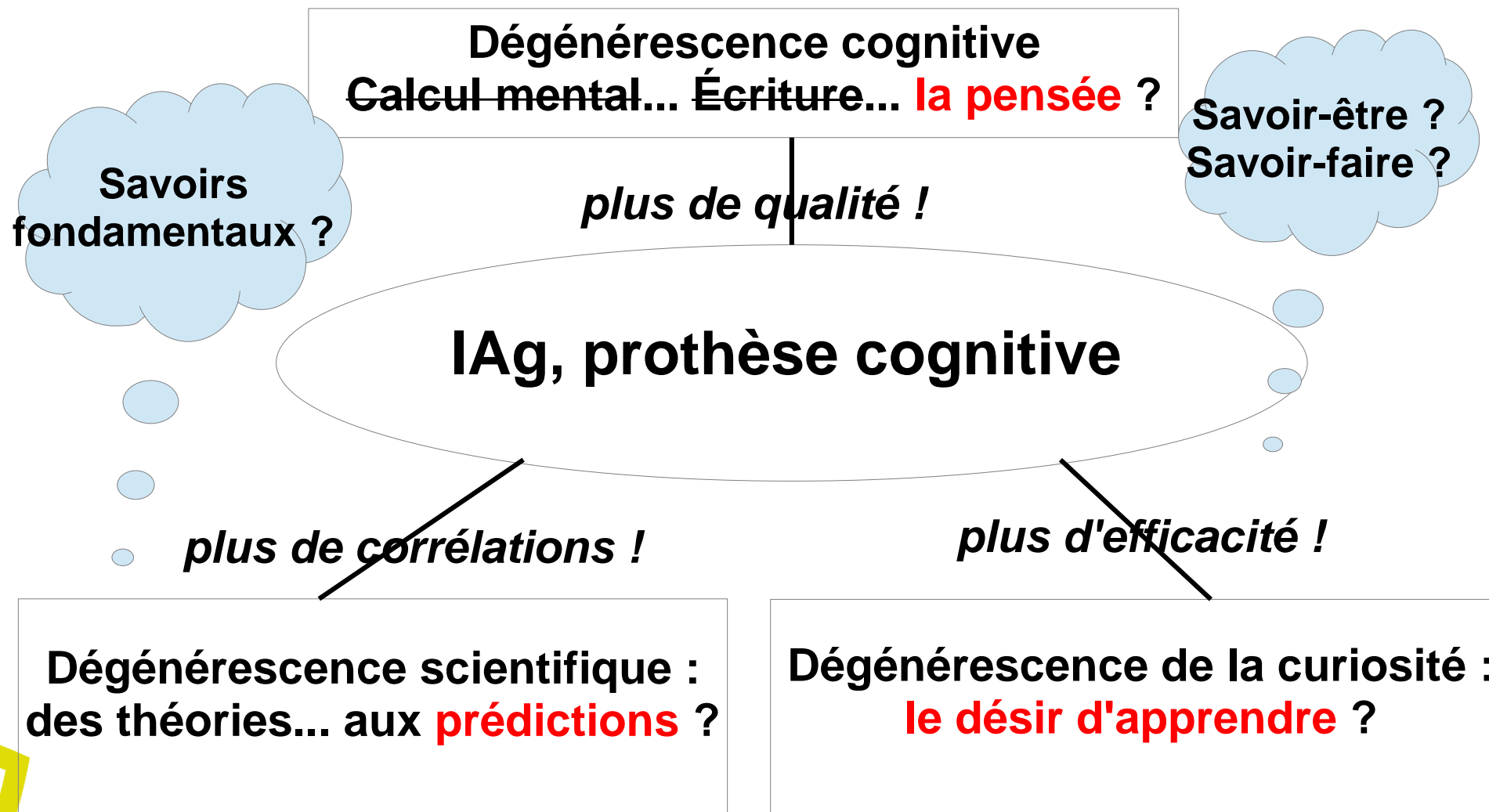
**Les « situations éthiques » :  
une pluralité d'impératifs, sur différentes boucles dialogiques**

Séminaire - Comment l'IA générative transforme nos métiers ?

18 octobre 2024

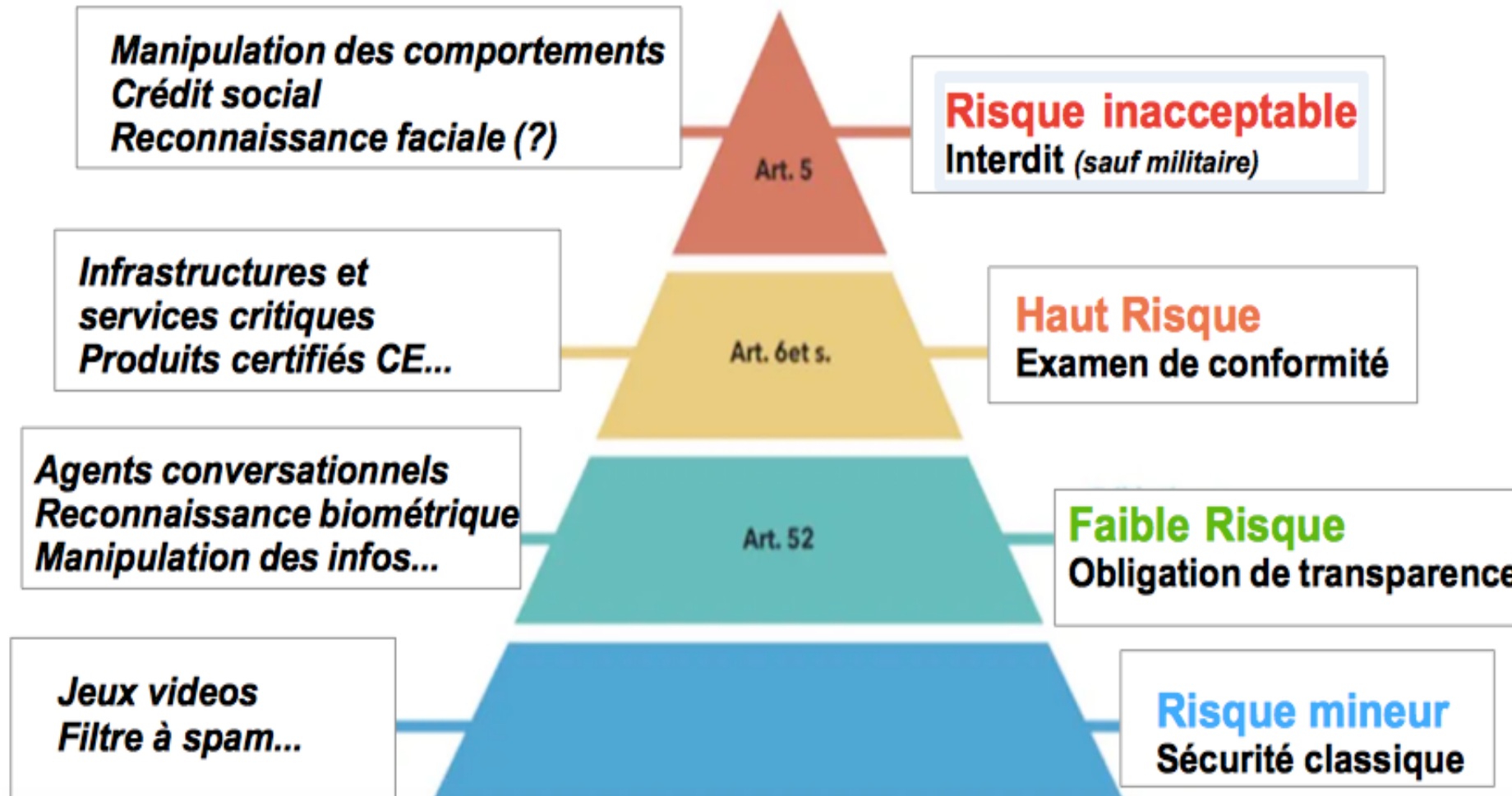
- b) « Un jour **une IA forte** dépassera l'intelligence humaine »
- - Une confusion sur **les intelligences** :
- **Intelligence prédictive** (performance : reconnaissance, probabilités, création) // **Intelligence humaine** (créativité : esprit critique et libre arbitre, en contexte socio, psy, biologique, spatial, temporel). Une **fracture** entre **élite consciente** // **crétin digital**
- - Une confusion sur **le rôle de la science** :
- **Prédiction** (probabilités, corrélations) // **Explication** (hypothèses, modélisation, théorisation, anticipation...)
- - Une confusion autour du « **transhumanisme** » :
- **La novlangue** (liberté, innovation, éternité... ) // **L'idéologie** (individualisme libertarien et/ou hiérarchie confucéenne)

# Du crétin digital (les algorithmes de recommandations)... au **crétin génératif** (la génération de contenus)



- c) « De toute façon, **on n'arrêtera pas l'innovation** »
- « *Celui qui croit que la croissance peut être infinie dans un monde fini est un fou ou ... ..* »
- - Un déni des **externalités** : **coût écologique** (extractions, énergie, pollutions), **coût social** (propagation des intoxications, addiction et propriété intellectuelle...), **perte de souveraineté** (santé, justice, école... espace, océans)
- - Une confusion entretenue entre **innovation et besoin du marché** : **le techno-solutionnisme** (monopolisation des choix) // **la gestion des biens communs** (choix institutionnel, E. Ostrom)
- - Une confusion entretenue sur **innovation et déréglementation** : **la techno-politique** (*Big Tech – Big states*) // **le droit** (AIAct, ~~SB-1047~~)

# La pyramide des risques de l'A.I.Act 2024



+ les « modèles à usage général » : Niveau 1 obligation de transparence et documentation, Niveau 2 mesures d'atténuation des risques

# Dialogue sur la sécurité de l'IA, Pékin 2024

## Aucun système d'IA ne devrait être capable de...

- **de se répliquer ou de s'améliorer** sans l'approbation et l'assistance explicites de l'homme ;
- d'augmenter considérablement la capacité des acteurs à **concevoir des armes de destruction massive**, ni violer la convention sur les armes biologiques ou chimiques ;
- **d'exécuter de manière autonome des cyberattaques** entraînant de graves pertes financières ;
- d'amener systématiquement ses concepteurs ou régulateurs **à mal comprendre sa capacité** à franchir l'une des lignes rouges précédentes.

# Merci pour votre attention

Des questions ?