

# Tout comprendre (ou presque) sur l'intelligence artificielle

Emilie Kaufmann (CNRS, CRISyAL)

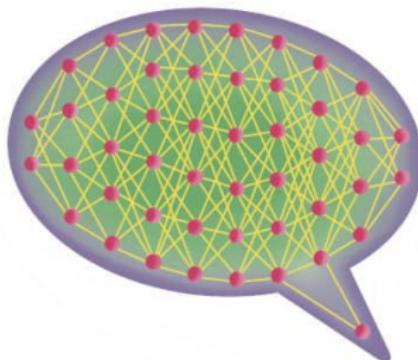


Semaine NSI, 10 décembre 2025

OLIVIER CAPPÉ

CLAIRE MARC

# TOUT COMPRENDRE (OU PRESQUE) SUR L'INTELLIGENCE ARTIFICIELLE

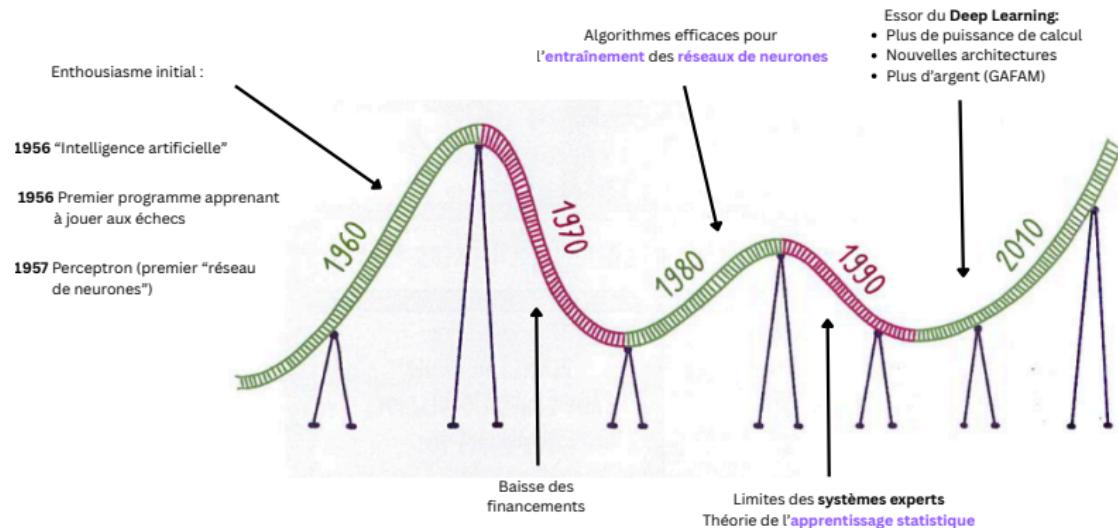


CNRS ÉDITIONS

# Qu'est-ce que l'Intelligence Artificielle?

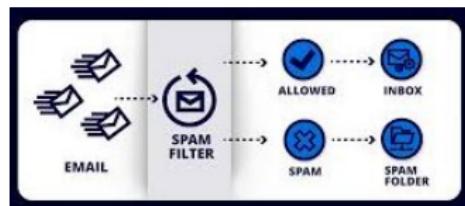
- IA  $\neq$  ChatGPT
  - IA = Ensemble des méthodes et des technologies permettant à des **systèmes d'IA** (comme ChatGPT) d'exister
- un **domaine de recherche**

# L'IA en quelques dates



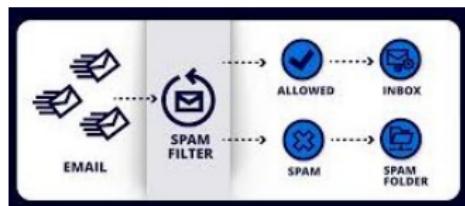
# Quelques tâches d'IA

- détection de spam



# Quelques tâches d'IA

- détection de spam



- reconnaissance d'écriture

Bonjour,  
je m'appelle  
Emilie

Bonjour, je m'appelle  
Émilie

# Quelques tâches d'IA

- classification d'image: patient sain / patient malade



# Quelques tâches d'IA

- classification d'image: patient sain / patient malade



- recommandation : quel film Netflix va vous proposer?

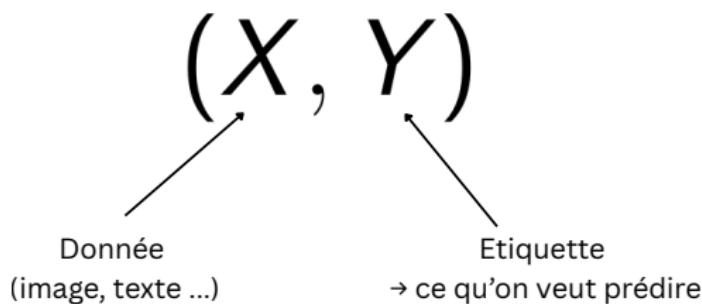


- 1 Une tâche d'IA: l'apprentissage supervisé
- 2 Un algorithme d'IA: l'apprentissage d'un réseau de neurones
- 3 Vers des systèmes d'IA complexes

- 1 Une tâche d'IA: l'apprentissage supervisé
- 2 Un algorithme d'IA: l'apprentissage d'un réseau de neurones
- 3 Vers des systèmes d'IA complexes

## Le problème

Etant donné un grand nombres de données annotées par des humains, arriver à **prédir l'étiquette d'une nouvelle donnée**.

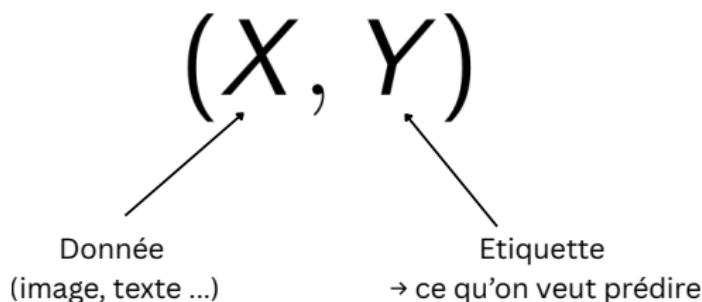


Exemple:

- $X$ : texte d'un e-mail
- $Y$ : spam / non-spam

## Le problème

Etant donné un grand nombres de données annotées par des humains, arriver à **prédir l'étiquette d'une nouvelle donnée**.



Exemple:

- $X$ : radio d'un patient (image)
- $Y$ : patient sain / patient malade

**Données d'entraînement:** de nombreux exemples bien étiquetés

$$(X_1, Y_1)$$

$$(X_2, Y_2)$$

...

$$(X_n, Y_n)$$

**Objectif:** Comprendre le lien entre  $X$  et  $Y$  en **généralisant** à partir de ces exemples

$$X \xrightarrow{\hat{f}_n} Y$$

- proposer un **prédicteur**  $\hat{f}_n$  (une *fonction*) tel que  $\hat{f}_n(X)$  soit proche de  $Y$ ... pour une nouvelle entrée  $(X, Y)$

## Exemple: Reconnaissance d'images



X: image d'un animal

- $X$  est représenté par les valeurs numériques des pixels de l'image (un vecteur), ici 100 nombres

Y: chien / pas chien

- $Y$  est représenté par une valeur dans  $\{0, 1\}$

# Un exemple de prédicteur

Un prédicteur *linéaire* attribue un **poids** (une valeur numérique ajustable) à chaque pixel.

ENSUITE, POUR CHAQUE PIXEL, LA VALEUR EST MULTIPLIÉE PAR LE POIDS :

$$\begin{matrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{matrix} \times \begin{matrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{matrix} = \begin{matrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{matrix}$$

ON ADDITIONNE TOUTS CES RÉSULTATS (LES PRODUITS PIXELS × POIDS)

$$= 13959$$

Si cette somme dépasse un certain seuil, le modèle décide que l'image représente un chien

CHIEN

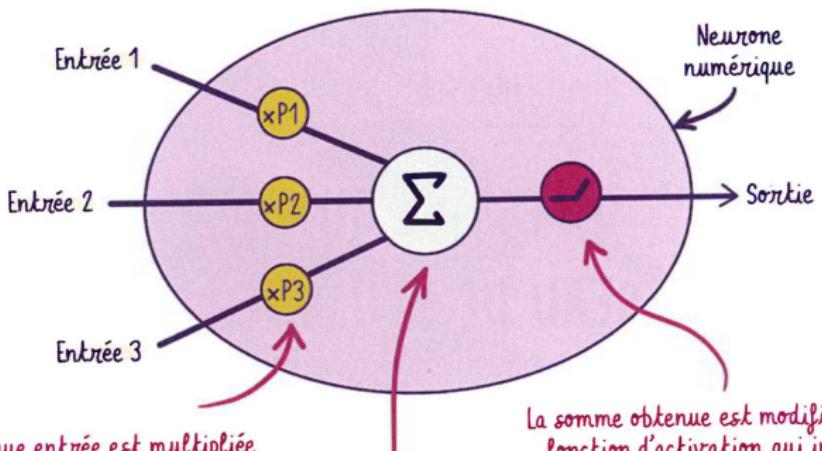
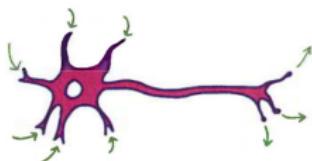
CHIEN

Si non, il considère l'image comme pas un chien

L'**apprentissage** consiste alors à ajuster les poids pour que le prédicteur fasse peu d'erreurs (sur de nouvelles images !).

- 1 Une tâche d'IA: l'apprentissage supervisé
- 2 Un algorithme d'IA: l'apprentissage d'un réseau de neurones
- 3 Vers des systèmes d'IA complexes

# Le neurone numérique

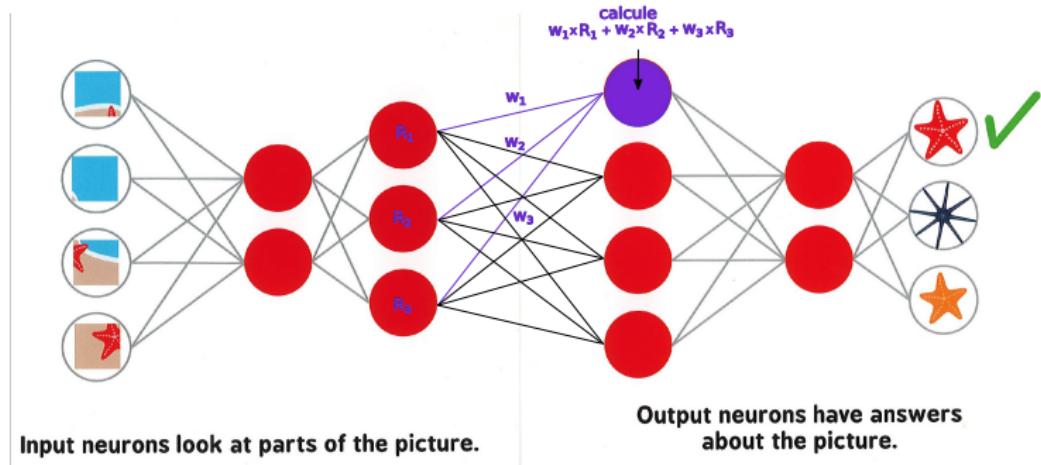


Chaque entrée est multipliée par un poids spécifique

Les produits sont ensuite additionnés

La somme obtenue est modifiée par une fonction d'activation, qui introduit généralement un effet de seuil

# Un réseau de neurones



En partant des neurones d'entrée (pixels de l'image), chaque neurones effectue les calculs précédents

(multiplication par les poids / somme / fonction d'activation)

→ Pour de grands réseaux, le calcul de la prédiction est déjà coûteux

# L'apprentissage: comment choisir les poids du réseau?

$w$  : ensemble des poids du réseaux

(un grand nombre de valeurs numériques)

$f_w$  : le réseau de neurones dont les poids sont  $w$

**Idée:** trouver les poids qui minimisent les erreurs sur la base d'apprentissage.

→ pour chaque exemple  $(X_i, Y_i)$  on veut que l'étiquette prédite  $f_w(X_i)$  soit proche de la vraie étiquette  $Y_i$

**Formalisation mathématique:** “bons poids”  $\leftrightarrow$  faibles valeurs d'une *fonction de perte*, par exemple

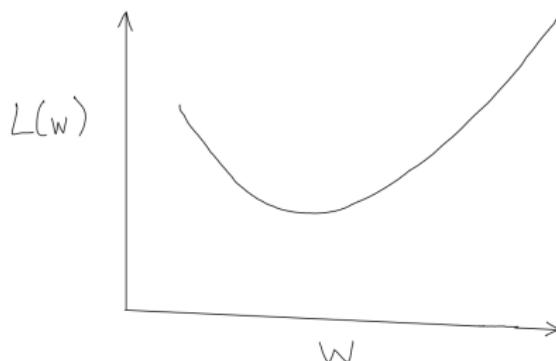
$$L(w) = \sum_{i=1}^n (Y_i - f_w(X_i))^2$$

# L'apprentissage: comment choisir les poids du réseau?

**But:** minimiser la fonction de perte

$$L(w) = \sum_{i=1}^n (Y_i - f_w(X_i))^2$$

- Calcul exact du minimiseur? Impossible.
- utilisation d'*algorithmes d'optimisation* pour s'en approcher

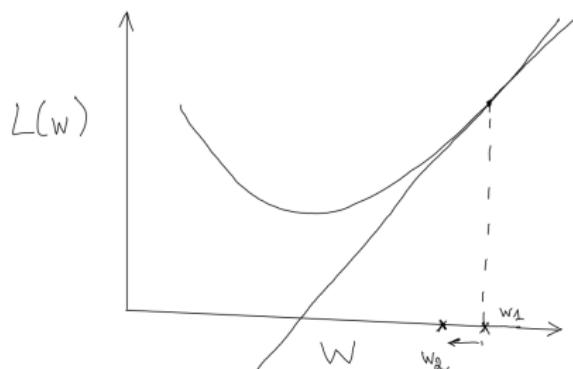


# L'apprentissage: comment choisir les poids du réseau?

**But:** minimiser la fonction de perte

$$L(w) = \sum_{i=1}^n (Y_i - f_w(X_i))^2$$

- Calcul exact du minimiseur? Impossible.
- utilisation d'*algorithmes d'optimisation* pour s'en approcher

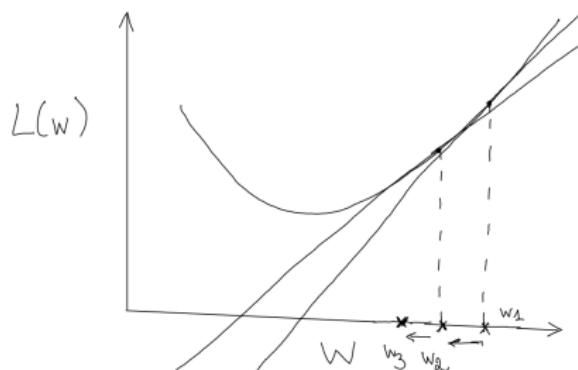


# L'apprentissage: comment choisir les poids du réseau?

**But:** minimiser la fonction de perte

$$L(w) = \sum_{i=1}^n (Y_i - f_w(X_i))^2$$

- Calcul exact du minimiseur? Impossible.
- utilisation d'*algorithmes d'optimisation* pour s'en approcher

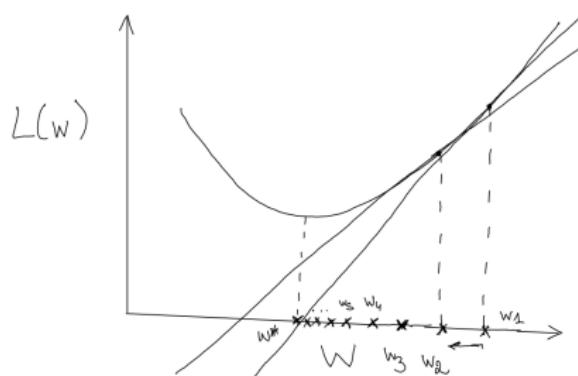


# L'apprentissage: comment choisir les poids du réseau?

**But:** minimiser la fonction de perte

$$L(w) = \sum_{i=1}^n (Y_i - f_w(X_i))^2$$

- Calcul exact du minimiseur? Impossible.
- utilisation d'*algorithmes d'optimisation* pour s'en approcher



# L'apprentissage: comment choisir les poids du réseau?

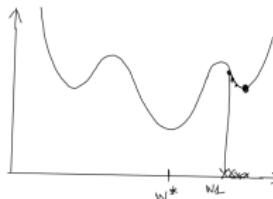
**But:** minimiser la fonction de perte

$$L(w) = \sum_{i=1}^n (Y_i - f_w(X_i))^2$$

- Calcul exact du minimiseur? Impossible.
- utilisation d'*algorithmes d'optimisation* pour s'en approcher

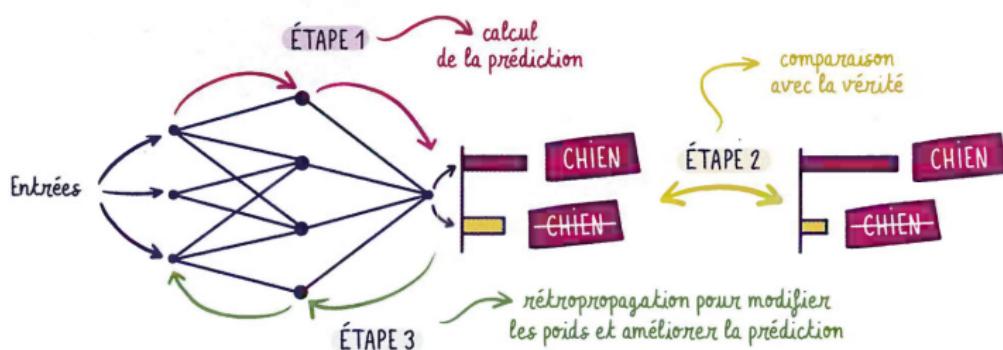
Difficulté:

- la fonction  $L$  est très complexe, on n'a pas vraiment de garantie d'atteindre le minimum...



# L'apprentissage: comment choisir les poids du réseau?

- Ces algorithmes nécessitent le calcul du *gradient* de  $L$
- Calculable efficacement pour un réseau de neurones en une passe inverse (algorithme dit de **rétro-propagation**)



# Qu'est-ce qu'un bon réseau de neurones?

**But:** mesurer le **pouvoir de généralisation** du prédicteur qui a été appris,  $\hat{f}_n = f_{\hat{W}_n}$

On a besoin de **nouvelles données étiquetées** (*données de test*)

$$(\tilde{X}_1, \tilde{Y}_1)$$

$$(\tilde{X}_2, \tilde{Y}_2)$$

...

$$(\tilde{X}_m, \tilde{Y}_m)$$

- calcul du pourcentage d'erreurs ( $\hat{f}_n(\tilde{X}_i) \neq \tilde{Y}_i$ ) sur ces nouvelles données (erreur de test)

- 1 Une tâche d'IA: l'apprentissage supervisé
- 2 Un algorithme d'IA: l'apprentissage d'un réseau de neurones
- 3 Vers des systèmes d'IA complexes

# Et ChatGPT dans tout ça?

**Brique de base:** apprentissage supervisé  
→ prédiction du mot suivant



**Alignement:** comment générer des “bonnes” réponses (correctes/éthiques/satisfaisante pour l’humain)

- collecte de **préférences humaines** (comparaisons de réponses) pour apprendre un modèle de récompense
- raffiner les poids de sorte à maximiser les récompenses (*apprentissage par renforcement*)

### ChatGPT : que signifie GPT

**G : Génératif**

—

Peut générer du texte



**P : Pré-entraîné**

—

A appris à partir d'un gros jeu de données



**T : Transformeur**

—

Architecture sur laquelle il repose



Transformer = le type de réseau de neurones utilisé

- mécanisme “d’attention”, révolution pour les modèles de langues (2017)
- nombre de poids :  $\simeq 100$  milliards pour GPT5
- données d’entraînement : données en libre accès sur Internet

# Et qu'en dit un système d'IA?

Je dois expliquer à des lycéens comment marche une IA comme toi ou ton copain ChatGPT, comment tu le ferais toi ?



## 2. Les 3 piliers d'une IA comme moi

### a) Les données : la "nourriture" de l'IA

- *"On m'a donné des milliards de textes (livres, articles, sites web) pour que je comprenne comment les mots s'assemblent, comment on pose des questions, comment on raconte une blague, etc."*

### b) L'apprentissage : le "cerveau" qui s'entraîne

- *"Grâce à des algorithmes (des recettes mathématiques très complexes), j'ai analysé ces textes pour repérer des motifs : après le mot 'bonjour', on dit souvent 'comment ça va ?', etc."*

### c) La prédiction : le "superpouvoir"

- *"Quand vous me posez une question, je ne cherche pas la réponse dans une base de données. Je prédis quel mot a le plus de chances de venir après le vôtre, comme quand vous envoyez un SMS et que votre téléphone propose le mot suivant."*

- L'intelligence artificielle: un **champ de recherches** comprenant de nombreuses questions
- De nombreuses briques doivent être combinées pour obtenir des **systèmes d'IA** performants
- L'entraînement de ces systèmes est très coûteux (énergétiquement), et leur utilisation aussi!

### 3. Les limites : ce que je ne sais pas faire

- "Je ne comprends pas vraiment ce que je dis (je ne ressens rien, je n'ai pas de conscience). Je ne fais que calculer des probabilités de mots !"
- "Je ne sais que ce qu'on m'a appris avant novembre 2024. Si vous me demandez qui a gagné la Coupe du Monde 2026, je ne saurai pas !"
- "Je peux faire des erreurs ou inventer des choses (on appelle ça des 'hallucinations'). Toujours vérifier mes réponses !"

Quelques directions de recherche qui nous intéressent :

- IA et équité (Michael Perrot, *Magnet*)
- IA et confidentialité (Thomas Michel, *Scool*)
- IA et explicabilité (Julie Jacques, *Orkad*)
- IA et sécurité (Jan Butora, *Sigma*)
- IA et consommation énergétique (Tristan Coignon, *Spirals*)