POKERGPT

Jouer (correctement) au poker avec un grand modèle de langage.

Narada MAUGIN, Tristan CAZENAVE
Université Paris-Cité, Université Paris Dauphine-PSL

QU'EST-CE QUE "LE" POKER ?

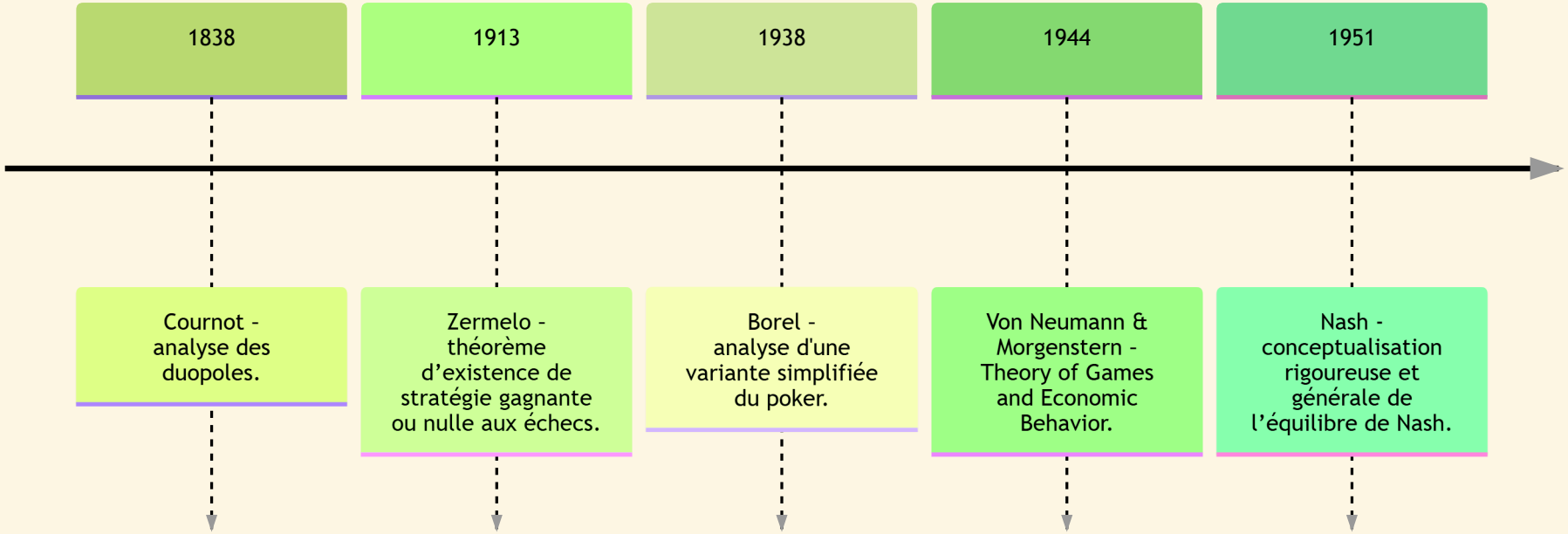
- Famille de jeux de cartes
  - variante principale, le **Texas Hold'em No-Limit**
- Apparu dans les années 1820 aux États-Unis
- ~ 100 millions de joueurs dans le monde
- Jeu à information incomplète
- Arbre de jeu immense

CITATION

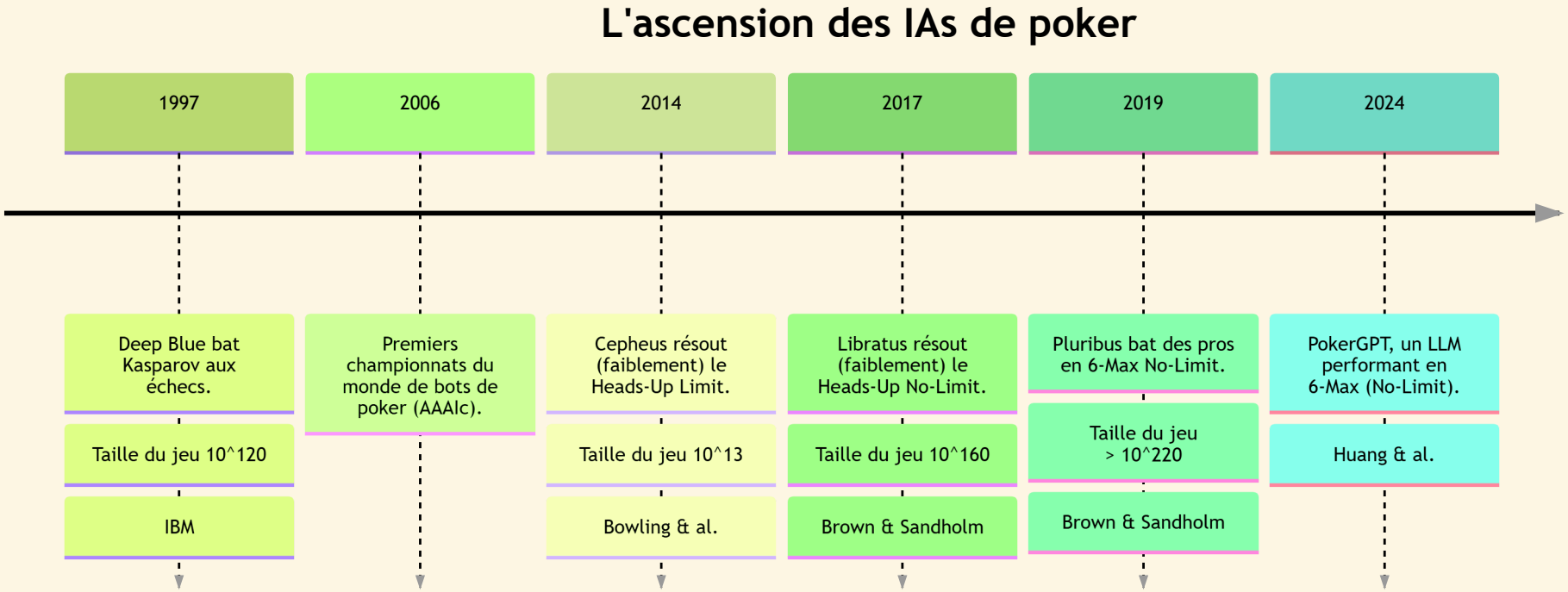
> « Les échecs sont une forme bien définie de calcul. Vous ne pouvez peut-être pas trouver les réponses, mais en théorie, il doit y avoir une solution […].
>
> Les vrais jeux ne sont pas du tout comme ça. La vraie vie n'est pas comme ça. **La vraie vie consiste à bluffer, à utiliser de petites tactiques de tromperie, à se demander ce que l'autre personne va penser que j'ai l'intention de faire.** Et c'est ce que sont les jeux dans ma théorie. »
>
> — John von Neumann (1903-1957)

FRISE CHRONOLOGIQUE (1)

Repères historiques de la théorie des jeux



FRISE CHRONOLOGIQUE (2)



QUELS DÉFIS RESTENT POUR LES IAS DE POKER ?

- Le défi principal : les tournois
  - Evolution du nombre de joueurs / de jetons
  - Les mains ne sont pas iid
  - Modèle *chipsEV* à revoir
- Les **spin&go**
  - Format le plus joué sur internet
  - Mini tournoi à 3 joueurs –> plus facile à aborder

POURQUOI UTILISER UN LLM AU POKER ?

- Les LLMs sont mauvais au poker en zero-shot et few-short prompting
  - Pas d'équivalent de GPT-3.5-turbo-instruct aux échecs
- Avantages :
  - Savoir préalable
  - Flexibilité
  - Raisonnements avancés
  - Prometteur

COLLECTE DES DONNÉES

- Mains jouées :
  - **Dates** : entre 2018 et 2020
  - **Site** : partypoker.fr
  - **Format** : spin&go
  - **Limites** : 50€, 100€ et 250€
  - n = 320 000

TRAITEMENT DES DONNÉES

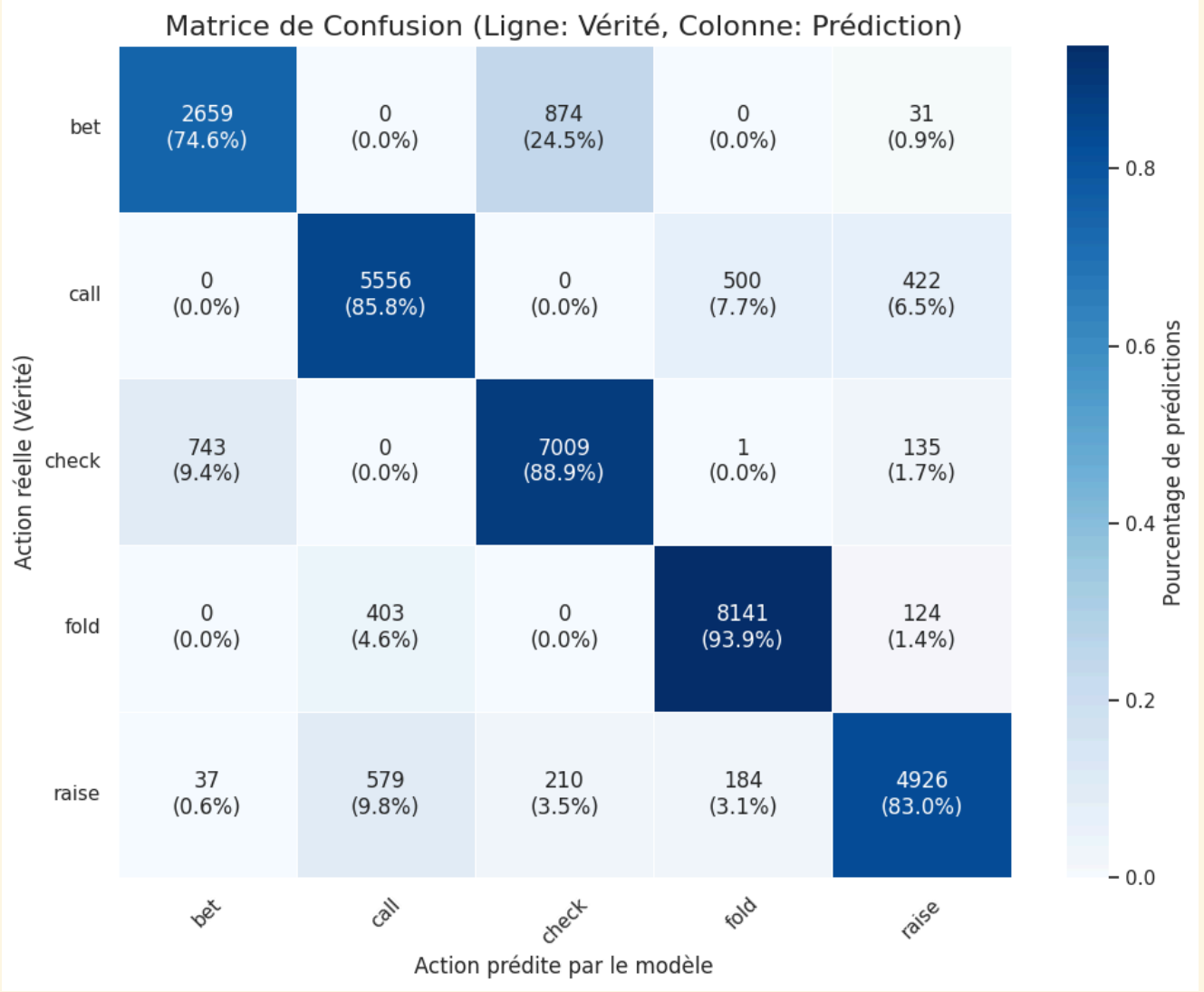
- Main traitée :
  {“instruction”:“pos:H=SB stacks:H=15.1,BB=22.4 hand:Kd7s | pre: H r2,BB c | flop:5h8d4s BB x,H x | turn:4d BB x,H x | river:Td SB x, H:””,“output”:“x”,“input”:““}
- Main originale :
  Game #20396217527 starts.
  Game #<do not remove this line!> starts.
  ***** Hand History for Game 20396217527 *****
  NL Texas Hold’em €250 EUR Buy-in - Friday, July 10, 18:41:29 CEST 2020
  Table 250€ SIT’N GO JAQKPOT (276572612) Table #1 (Real Money)
  Seat 2 is the button
  Total number of players : 2/3
  Seat 1: Dimitrov98 (605)
  Seat 2: FrenchBaguette (895)
  Trny: 276572612 Level: 3
  Blinds(20/40)
  FrenchBaguette posts small blind [20].

FINE-TUNING SUPERVISÉ (SFT)

- **Modèle** : Llama 3.1-8B-Instruct
- **Avantages** :
  - Disponible facilement
  - Relativement petit
- **Désavantage** :
  - Faible niveau au poker
- **Temps d'entraînement** : 10 heures
- **Paramètres** :
  - LoRA (~10^7 paramètres)
  - cutoff_len = 128
  - epoch = 4
  - learning_rate = 5e-5

MATRICE DE CONFUSION

- Testing set (n = 32 534)
- Précision (*exact accuracy*):
  - 80%
  - 84% (avec tolérance)
- Coups illégaux = 1,4%
- F1-score = 85,4%



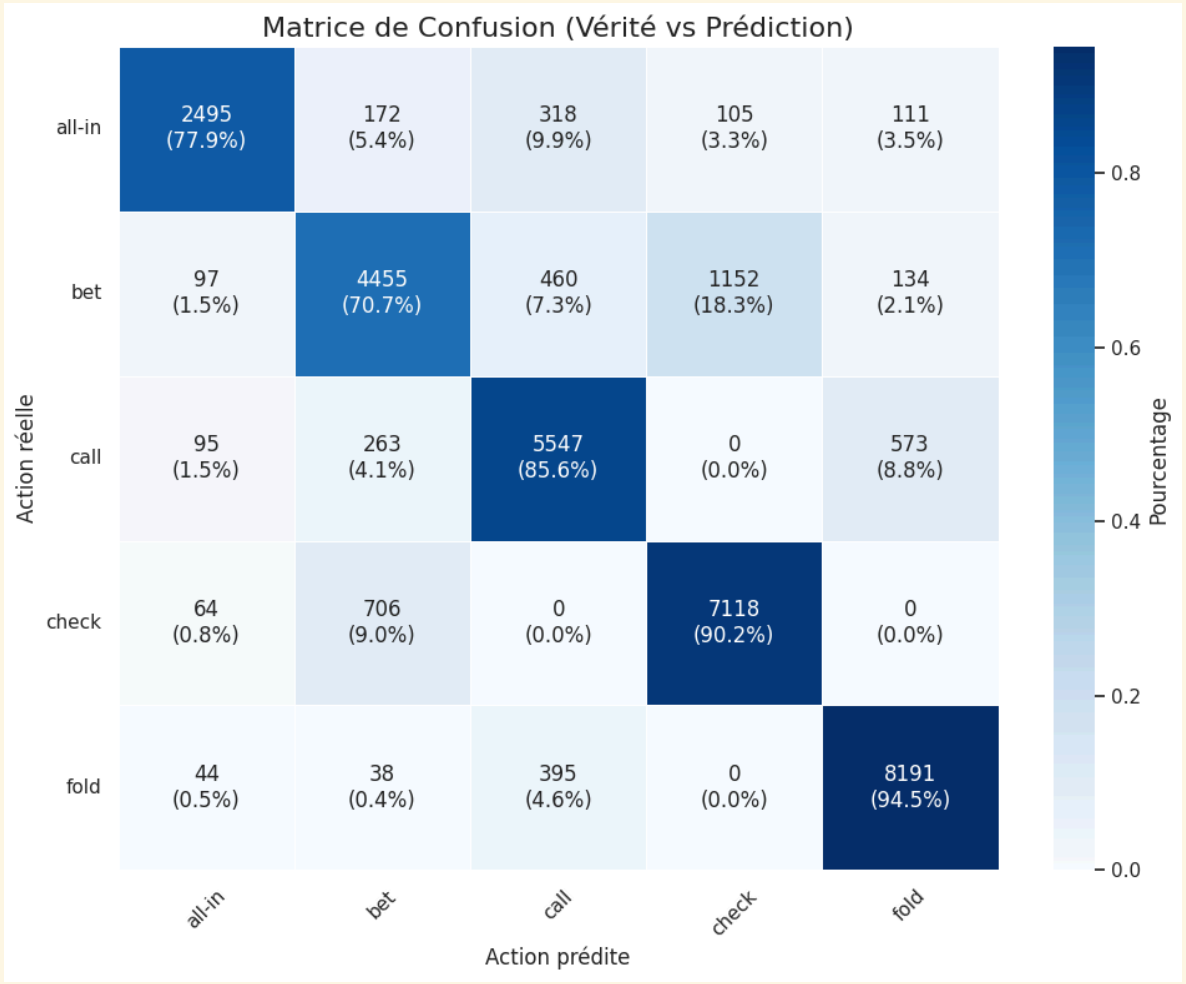
ÉVALUATION

- Évaluer l'IA ?
- Contre Slumbot, vainqueur de l'AAAIc 2018
  - API accessible
  - Bon niveau
- Perte colossale
- Correctif, empêcher les all-in en les remplaçant par :
  - une mise 2/3 du pot si l'adversaire n'a pas misé
  - une relance à hauteur de 3 fois la mise adverse sinon

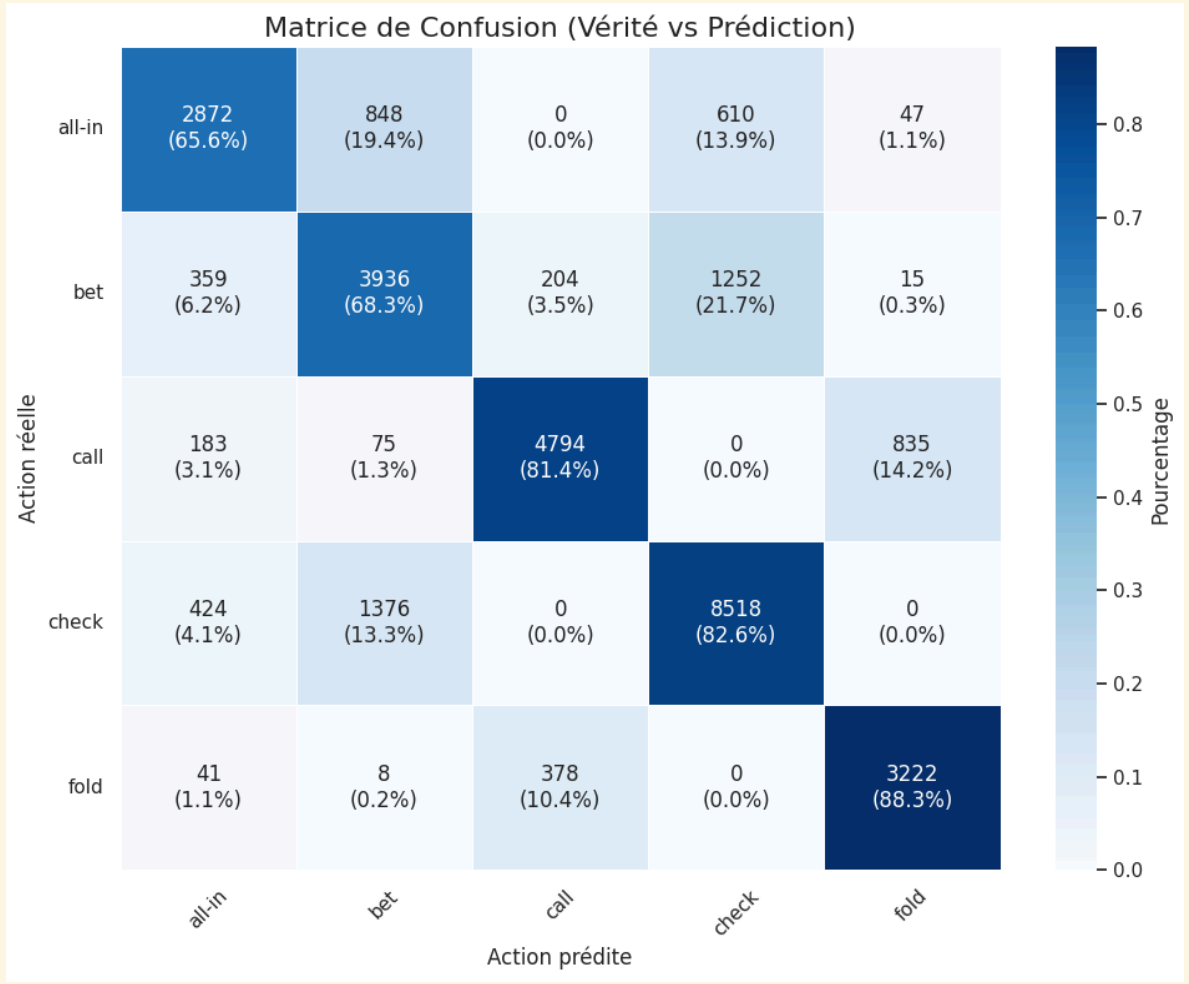
- Résultats contre Slumbot (en BB/100) :
  - Notre IA sur 30 000 mains :
    - **13.4 ± 12.9** (IC 95 %)
    - Temps d'action : 0,4 sec
  - ReBel (2020)
    - **4.5 ± 0.5** (IC 68%)
  - AlphaHoldem (2022)
    - **11.1 ± 1.61** (IC 95 %)
  - PokerGPT (2024)
    - **15.8 ± 4.9** (IC 95 %)

REINFORCEMENT LEARNING (RL)

- Objectif : se rapprocher du GTO
- Utilisation d'InstaGTO
- RL offline en utilisant ORPO
- Entraînement sur 320 000 mains
  - 270 000 mains d'InstaGTO (données synthétiques)
  - 50 000 mains du pro (pour éviter le *catastrophic forgetting*)



ORPO sur le dataset de test pro (n = 32 534). Précision : 78.6% (83.2% avec tolérance.)



ORPO sur le dataset de test solver (n = 30 000). Précision : 71.9% (77.8% avec tolérance).

ÉVALUATION

- Match entre les modèles SFT et SFT+ORPO
  - 5 000 mains (format duplicate)
  - **14.1 ± 10.2** BB/100 (IC 95%) pour SFT+ORPO
  - Le RL a fonctionné
- Limites
  - Pas de tests en conditions réelles à grande échelle
  - Quantité incomprise (5 < 5.2 < 5.11 ?!)

AMÉLIORATIONS POSSIBLES

- Meilleur modèle fondation
- Données de meilleures qualités
- RL sur mains réelles
- Adaptation à son adversaire
- …

CONCLUSION ET PERSPECTIVES

- Ça marche !
- Pour appronfondir :
  - Peut-on détecter le bluff chez un LLM ?
  - Le LLM peut-il expliquer honnêtement pourquoi il joue tel coup ?
  - Existe-t-il des prompts qui le font déjouer ?
  - Comment détecter une IA de poker ?
- Défiez notre IA : bit.ly/pokerpfia !
- Des questions ?