

# Rapport de Stage

«Le Rôle de l'IA dans les Comportements  
Stratégiques en Économie Industrielle»  
Au titre d'une contribution à un projet de  
recherche sur l'apport de l'IA dans les stratégies  
de prix des acteurs économiques

L2 MIASHS 2024–2025

Mohamed HAMLIL

*Sous la supervision de Alexis Garapin (Professeur UGA)  
et d'Olivier Bonroy (Directeur de recherches INRAE)*

Laboratoire GAEL

Stage du 23 janvier 2025 au 14 avril 2025



# Table des matières

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Aperçu général du laboratoire . . . . .	2
1.2	Contexte du stage . . . . .	2
1.3	Consolidation des compétences . . . . .	2
<b>2</b>	<b>État de l’art</b>	<b>3</b>
2.1	Cadre conceptuel . . . . .	3
2.2	Méthodologie de sélection . . . . .	4
2.3	Analyse des contributions . . . . .	4
2.3.1	Decoding GPT’s Hidden Rationality of Cooperation . . . . .	4
2.3.2	Ultimatum Bargaining : Algorithms vs. Humans . . . . .	5
2.3.3	When communicative AIs are cooperative actors : A prisoner’s dilemma experiment on human–communicative artificial intelligence	5
2.4	Synthèse et discussion . . . . .	5
<b>3</b>	<b>Travail supplémentaire</b>	<b>6</b>
3.1	Lectures complémentaires . . . . .	6
3.2	Projets réalisés . . . . .	7
<b>4</b>	<b>Conclusion</b>	<b>7</b>

# 1 Introduction

## 1.1 Aperçu général du laboratoire

Le Laboratoire d'Économie Appliquée de Grenoble ([GAEL](#)) est une unité mixte de recherche affiliée au CNRS, à l'INRAE, à l'Université Grenoble Alpes et à Grenoble INP. Ses domaines de recherche portent principalement sur l'innovation et la consommation durables dans les secteurs énergétiques et agro-industriels. Les méthodologies clés du laboratoire incluent la modélisation microéconomique (économie industrielle, économie comportementale), l'économie expérimentale, l'économétrie, la modélisation appliquée à certains secteurs (notamment l'énergie) et l'analyse qualitative fondée sur des études de cas.

## 1.2 Contexte du stage

Dans le cadre du programme de l'Université Grenoble Alpes et de Grenoble INP, la mise en situation professionnelle permet à l'étudiant de deuxième année de développer des compétences techniques et un savoir-être professionnel. C'est dans ce contexte que j'ai intégré, pour un stage de douze semaines (23 janvier – 14 avril 2025) au sein du laboratoire GAEL, un projet de recherche explorant le rôle de l'IA dans les comportements stratégiques en économie industrielle. Ce travail a été mené sous l'encadrement de M. Alexis Garapin (Professeur à l'UGA) et de M. Olivier Bonroy (Directeur de recherche à l'INRAE). Le projet combine une modélisation théorique et des expériences en laboratoire confrontant des participants humains à des algorithmes de fixation des prix, dans le but d'évaluer si l'IA peut favoriser des comportements coopératifs (par exemple, dans un dilemme du prisonnier).

## 1.3 Consolidation des compétences

Durant ce stage, j'ai pu mesurer l'importance des recherches menées par GAEL sur l'innovation et la consommation durables dans divers secteurs. Les méthodologies employées influencent les stratégies économiques globales et comprennent : la modélisation microéconomique, l'économie expérimentale, l'économétrie et l'analyse qualitative basée sur des études de cas. Afin de mener à bien ma mission, une fiche d'activités a été élaborée et portait sur les points suivants :

1. la rédaction d'une revue de littérature à partir d'articles académiques sélectionnés ;
2. le développement d'un algorithme pour les expériences en laboratoire.

Le laboratoire m'a mis à disposition un ensemble de ressources, notamment l'accès aux codes sources et un bureau dédié, pour réaliser ces travaux.

## 2 État de l’art

La coopération est au cœur de nombreuses interactions sociales et économiques, des dilemmes du prisonnier aux négociations d’ultimatum. Avec l’essor des systèmes d’intelligence artificielle (IA) capables d’échanger, de raisonner et d’apprendre, se pose la question : **l’IA peut-elle favoriser des comportements coopératifs ?** Cet état de l’art examine trois contributions majeures qui explorent les dynamiques coopératives impliquant des IA :

- *Decoding GPT’s Hidden Rationality of Cooperation* (Bauer *et al.*, 2023) BAUER *et al.* 2023,
- *Ultimatum Bargaining : Algorithms vs. Humans* (Özkes *et al.*, 2024) OZKES *et al.* 2024,
- *When communicative AIs are cooperative actors : A prisoner’s dilemma experiment on human-communicative artificial intelligence* (Yu-Leung Ng, 2023) NG 2023.

L’objectif est de [synthétiser](#) leurs méthodologies et résultats afin d’évaluer dans quelles conditions l’IA peut réellement soutenir ou stimuler la coopération.

### 2.1 Cadre conceptuel

1. **Coopération** Stratégie par laquelle un ou plusieurs acteurs acceptent un coût ou un risque immédiat afin de maximiser ultérieurement un bénéfice collectif ou mutuel, souvent analysée via les notions de réciprocité (tit-for-tat), d’altruisme et de confiance.
2. **Dilemme du prisonnier** Jeu fondamental de la théorie des jeux illustrant le conflit entre intérêt individuel et gain collectif, décliné en trois variantes :
  - (a) *Statique (one-shot)* : Les deux joueurs choisissent simultanément une action (coopérer ou trahir) pour un seul tour ; l’équilibre de Nash conduit généralement à la défection mutuelle.
  - (b) *Séquentiel* : Les décisions sont prises à tour de rôle ; le joueur qui agit en second peut conditionner son choix à l’observation de l’action du premier (Stackelberg), ce qui peut favoriser la coopération sous certaines conditions.
  - (c) *Répété (itéré)* : Le même dilemme se joue plusieurs fois, avec possibilité de mémoire des tours précédents ; stratégies de réciprocité (tit-for-tat, grim trigger) permettent de soutenir la coopération sur le long terme.
3. **Jeu de l’ultimatum** Interaction simultanée où un proposant propose une répartition d’une ressource (somme d’argent, points, etc.) et un répondant accepte ou rejette l’offre. La justice perçue se mesure au niveau des offres minimales acceptables et des seuils de sanction, révélant les normes sociales d’équité.
4. **Intelligence artificielle communicative** Agents basés sur des modèles de langage (LLM, chatbots) capables d’émettre et de recevoir du « cheap talk » (messages non contraignants), d’expliquer leurs décisions (explicabilité) et d’influencer la perception ou le comportement d’un interlocuteur humain ou algorithmique.

5. **Cheap talk** Communication non contraignante entre joueurs (messages avant ou pendant le jeu) qui peut faciliter la coopération en établissant des promesses, en signalant des intentions ou en modulant les attentes réciproques.
6. **Explicabilité** Degré et qualité des justifications fournies par un agent (humain ou IA) pour ses décisions : va de l'absence d'information à une explication détaillée, impactant la confiance et la perception d'équité.
7. **Équilibre de Nash bayésien** Solution du dilemme où chaque joueur, connaissant ses propres croyances probabilistes sur l'autre, choisit une stratégie optimisant son espérance de gain, prenant en compte information asymétrique et incertitude.
8. **Normes sociales et justice procédurale** Principes culturels internes (fair-play, réciprocité, sanction des injustices) qui façonnent les seuils d'acceptation et les réactions aux propositions perçues comme inéquitables.

## Mots-clés

Coopération – Théorie des jeux – Réciprocité – Tit-for-tat – Dilemme du prisonnier – Jeu de l'ultimatum – Cheap talk – Explicabilité – Équilibre de Nash – Normes sociales.

## 2.2 Méthodologie de sélection

Les trois travaux retenus répondent à ces critères :

- **Interaction directe avec des IA** (modèles de langage, algorithmes de négociation, agents communicants).
- **Comparaison IA vs humain** dans des contextes de décision stratégique.
- **Approche expérimentale** avec données comportementales sur de larges échantillons.

## 2.3 Analyse des contributions

### 2.3.1 Decoding GPT's Hidden Rationality of Cooperation

- **Auteurs** : Bauer, Liebich, Hinz & Kosfeld (2023).
- **Protocole** : confrontation de GPT-3.5 et GPT-4 à un dilemme du prisonnier séquentiel face à des adversaires humains et simulés.
- **Objectif** : mesurer dans quelle mesure ces LLMs imitent la rationalité et la coopération humaines.
- **Résultats clés** :
  - *Comportement comparé des GPT et des humains* Rôle de second joueur (sans incertitude) : GPT coopère nettement plus souvent que les humains, surtout si le premier joueur a défecté. Rôle de premier joueur (sous incertitude) : GPT-3.5

coopère moins que les humains quand la réponse de l'autre est inconnue, tandis que GPT-4 coopère davantage.

- *Croyances optimistes des LLM* Les deux versions estiment systématiquement plus haut la probabilité de recevoir une coopération d'un adversaire humain, quel que soit leur propre choix initial.
- *Modèles de rationalité économique* Le modèle « homo œconomicus » n'explique quasiment pas le comportement des GPT. Le modèle « welfare conditionnel » rend compte d'environ 84,5

### 2.3.2 Ultimatum Bargaining : Algorithms vs. Humans

- **Auteurs** : Özkes, Hanaki, Vanderelst & Willems (2024).
- **Protocole** : jeu de l'ultimatum contre des partenaires humains ou algorithmiques (imitation, optimisation).
- **Échantillon** : 404 participants britanniques.
- **Résultats clés** :
  - *Seuil minimal d'acceptation d'offre (MAO)* Aucune différence significative de MAO selon que le partenaire est humain ou algorithmique, ni selon le niveau d'explicabilité de l'algorithme.
  - *Offres des proposants* Les partages 50–50 restent majoritaires, surtout face à un humain. Face à un algorithme, les participants offrent en moyenne 4 points de moins.
  - *Préférences des joueurs* Après le jeu, 40

### 2.3.3 When communicative AIs are cooperative actors : A prisoner's dilemma experiment on human–communicative artificial intelligence

- **Auteur** : Yu-Leung Ng (2023).
- **Protocole** : expérience itérée du dilemme du prisonnier, comparant participants face à partenaires humains et IA, avec ou sans communications normatives.
- **Résultats clés** :
  - La stratégie du partenaire (coopérative vs non coopérative) influence fortement le taux de coopération.
  - Les participants ne différencient pas le type de partenaire (IA ou humain) dans leur propension à coopérer.
  - Émergence d'une stratégie de *tit-for-tat* effective.

## 2.4 Synthèse et discussion

Les trois études proposent des éclairages complémentaires :

1. **Modélisation de la rationalité coopérative** Bauer *et al.* dévoilent un « welfare conditionnel » dans GPT, surpassant parfois la coopération humaine.
2. **Normes sociales et équité** Özkes *et al.* confirment que les normes d'équité guidant les humains s'expriment aussi face aux algorithmes.
3. **Communication stratégique** Ng montre que le *cheap talk* et les justifications normatives stabilisent la coopération sur le long terme.

**Synergie des approches** Le modèle économique de GPT fournit un cadre de références, validé empiriquement dans le jeu de l'ultimatum, puis renforcé par la communication normative.

**Potentiel coopératif de l'IA** En combinant modélisation, normes sociales et capacités de dialogue, l'IA peut non seulement imiter mais aussi stimuler des dynamiques coopératives robustes.

**Limites et risques** Sans supervision et transparence, les IA restent sujettes à la défection et à des croyances irréalistes, mettant en péril la confiance.

### Recommandations pour la conception

- Intégrer des objectifs collectifs dans les fonctions de récompense.
- Garantir une explicabilité fine des décisions.
- Proposer des interfaces de communication riches pour médiatiser le *cheap talk* et les justifications normatives.

## 3 Travail supplémentaire

### 3.1 Lectures complémentaires

Pour approfondir le sujet, j'ai consulté plusieurs [articles](#) articles supplémentaires :

- Sandholm (1996) : Q-learning et convergence vers des équilibres variés (SANDHOLM et CRITES [1996](#)).
- Wang (2018) : influence de la mémoire et de la séquentialité dans le dilemme du prisonnier (WANG et al. [2018](#)).
- Horton (2023) : théorie de l'« Homo Silicus » in silico (HORTON [2023](#)).

### 3.2 Projets réalisés

J’ai mis en œuvre un protocole expérimental (sélection des agents, paramétrisation et analyse) hébergé dans un dépôt [GitHub](#) dédié, comprenant :

- **Étude des neurones miroirs** Analyse des comportements mimétiques (*Tit-for-Tat*) dans le projet [Neurones\\_Miroirs](#).
- **Agent stratégique en Prolog** Conception et implémentation d’un algorithme fondé sur les [consignes du jeu Prolog](#), pour simuler et comparer diverses stratégies .

## 4 Conclusion

Ce stage de douze semaines au laboratoire GAEL m’a offert l’opportunité de :

- Approfondir ma compréhension des branches de l’économie appliquée – industrielle, comportementale et expérimentale – en confrontant modèles théoriques et expérimentations en laboratoire.
- M’initier à la construction de modèles économiques et à l’élaboration de questions scientifiques, notamment grâce aux séminaires et aux échanges réguliers avec l’équipe GAEL.
- Découvrir les enjeux récents de l’IA appliquée à l’économie, via la synthèse d’articles académiques sur les LLM comme agents économiques, l’apprentissage par renforcement multi-agent, les biais et les hallucinations, ainsi que le concept d’« Homo Silicus ».
- Acquérir une démarche scientifique rigoureuse : rédaction de README détaillés, structuration de dépôts Git, documentation de notebooks Jupyter et adoption de bonnes pratiques de recherche reproductible.
- Développer mes compétences techniques : programmation d’agents stratégiques en Prolog (tournoi, convergence vers l’équilibre) et en Python (notebooks d’analyse, visualisations `pgfplots`).
- Mettre en œuvre des méthodes d’analyse statistique (ANOVA, tests de comparaison) pour interpréter les taux de coopération et évaluer la robustesse des simulations.
- Maîtriser l’usage de LaTeX pour la rédaction scientifique et de Beamer pour la préparation de la soutenance, renforçant mes capacités de communication écrite et orale.
- Appliquer concrètement des concepts de théorie des jeux (*Tit-for-Tat*, dilemme du prisonnier itéré) et d’apprentissage automatique, en développant une boîte à outils modulaire pour prototyper et comparer différentes stratégies.

Ce stage a renforcé mon intérêt pour l’économie et constitue une base solide pour mes projets futurs, notamment ma candidature à la bourse d’excellence MIAI.



## Références

- BAUER, Kevin et al. (2023). “Decoding GPT’s hidden “rationality” of cooperation”. In : *SAFE Working Paper No. 401*. URL : <https://ssrn.com/abstract=4576036>.
- HORTON, John J. (2023). *Large Language Models as Simulated Economic Agents : What Can We Learn from Homo Silicus ?* Rapp. tech. NBER Working Paper No. 31122. National Bureau of Economic Research. URL : <http://www.nber.org/papers/w31122>.
- NG, Yu-Leung (oct. 2023). “When communicative AIs are cooperative actors : A prisoner’s dilemma experiment on human–communicative artificial intelligence”. In : *Behaviour and Information Technology* 42.13, p. 2141-2151. DOI : [10.1080/0144929X.2022.2111273](https://doi.org/10.1080/0144929X.2022.2111273). URL : <https://doi.org/10.1080/0144929X.2022.2111273>.
- OZKES, Ali I. et al. (2024). “Ultimatum bargaining : Algorithms vs. Humans”. In : *Economics Letters* 244, p. 111979. DOI : [10.1016/j.econlet.2024.111979](https://doi.org/10.1016/j.econlet.2024.111979).
- SANDHOLM, Tuomas W. et Robert H. CRITES (1996). *Multiagent Reinforcement Learning in the Iterated Prisoner’s Dilemma*. Rapp. tech. UM-CS-1996-015. Supported by AFOSR Grant F49620-93-1-0450. University of Massachusetts Technical Report.
- WANG, Weixun et al. (2018). “Towards cooperation in sequential prisoner’s dilemmas : A deep multiagent reinforcement learning approach”. In : *arXiv preprint arXiv :1803.00162*.