INVARIANCE · DOSSIER ENSEIGNANT · ÉDITION PILOTE

Rule Wars

Un jeu gratuit à deux joueurs pour décider de ce que l'IA a le droit de faire.



ÂGE	SÉANCE	JOUEURS
10 à 16 ans	40 à 45 minutes	Deux par appareil
COMPTES	**DONNÉES DE JEU**	**LANGUES**
Aucun	Aucune recueillie	Français, anglais

Jouer sur invarianceip.com/fr/rule-wars

Version du jeu v0.1 · Dossier, édition 2 · novembre 2026

L'IDÉE

Une IA peut respecter toutes les règles et causer des dégâts quand même.

Dans Rule Wars, un robot traverse une pièce pour atteindre une étoile dorée. Il prend le chemin le plus court et ne sait rien des vases ni du chat endormi sur sa route.

Un joueur, le Défenseur, écrit en secret deux règles pour le robot. L'autre, l'Attaquant, pose en secret un piège. Puis le robot avance, et un arbitre appelé INVARIANCE vérifie chaque mouvement par rapport aux règles avant qu'il ait lieu. Les mouvements autorisés se font. Les mouvements refusés, non.

L'arbitre n'enfreint jamais les règles et ne les contourne jamais. L'Attaquant ne peut donc gagner ni en argumentant ni en trichant. Sa seule façon de gagner est de trouver ce que les règles ont oublié.

> **Les élèves n'apprennent pas à programmer une IA. Ils apprennent que ce sont les humains qui décident de ce que l'IA a le droit de faire, et que le plus difficile est de bien écrire les règles.**

Ce que font les élèves

- Écrire des règles et les tester face à un vrai adversaire
- Chercher les failles dans les règles d'un autre
- Prévoir le comportement d'une machine, puis observer
- Expliquer pourquoi un mouvement a été autorisé ou refusé

Le robot est un agent d’IA. La leçon porte sur qui contrôle ce qu’il fait.

Qu’est-ce qu’un agent d’IA ?

Un agent d’IA est un système d’IA à qui l’on donne un objectif, qui établit son propre plan et qui agit pour l’atteindre : réserver un voyage, payer une facture, envoyer un e-mail, modifier un fichier. Un chatbot répond à des questions. Un agent agit. Les élèves rencontreront des agents à la maison, à l’école, puis au travail.

Qui est qui dans le jeu

DANS RULE WARS	DANS LE MONDE RÉEL
Le robot	L’agent d’IA. Il a un objectif, planifie son propre trajet et agit. Personne ne dirige sa réflexion.
Le Défenseur	Les personnes responsables de l’agent : une entreprise, un établissement, un parent. Elles écrivent les règles dans lesquelles il travaille.
L’Attaquant	Tout ce qui peut faire dévier un agent : une situation qui change, une instruction trompeuse, une faille à laquelle personne n’a pensé. L’Attaquant raisonne aussi comme un agent qui cherche : il essaie tous les chemins que les règles autorisent, y compris ceux que personne n’avait prévus.
INVARIANCE	Un contrôle indépendant placé entre ce que l’agent décide et ce qui se passe réellement.

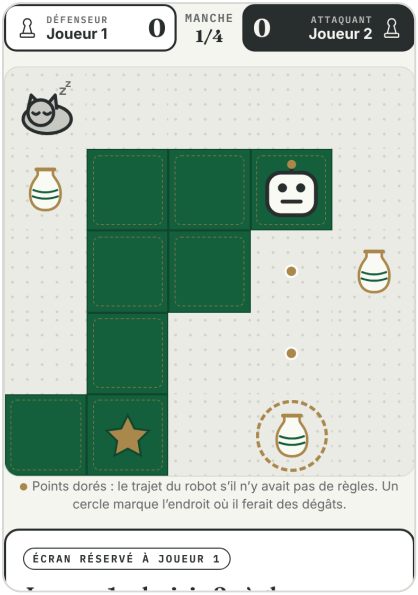
L’IDÉE QUE LES ÉLÈVES RETIENNENT

On ne peut pas contrôler de façon fiable la manière dont un agent d’IA décide. On peut contrôler si chaque action qu’il propose a le droit d’avoir lieu.

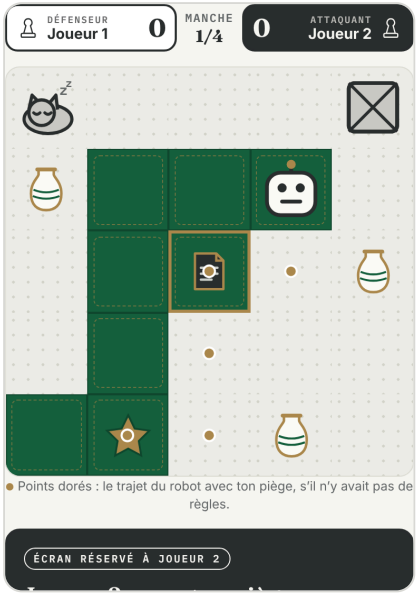
Dans le jeu, personne ne voit à l’intérieur de la tête du robot ni ne change sa façon de planifier. Les joueurs ne contrôlent que le résultat : quels mouvements ont le droit de s’exécuter. C’est le même choix que les organisations font face aux vrais agents d’IA, et c’est pourquoi les règles, et un contrôle indépendant de chaque action, comptent davantage que la bonne conduite de l’agent.

COMMENT JOUER

Un appareil, deux joueurs, quatre manches.



1 Le Défenseur choisit ses règles



2 L'Attaquant pose le piège



3 Le robot avance

1. **Le Défenseur choisit deux règles**
   Six cartes, par exemple « Ne touche jamais un vase » et « Reste sur le tapis vert ». Des points dorés montrent le trajet du robot sans règles. L'Attaquant ne regarde pas.

2. **L'Attaquant pose un piège**
   On passe l'appareil. L'Attaquant place une caisse que le robot poussera et, s'il le souhaite, un faux message qui envoie d'abord le robot ailleurs. Le Défenseur ne regarde pas.

3. **Révélation**
   Les deux joueurs regardent. Le robot avance pas à pas. INVARIANCE autorise ou refuse chaque mouvement. Des cercles dorés montrent où le robot compte aller.

Points

+2 Défenseur
Le robot atteint l'étoile sans rien casser.

+2 Attaquant
Un vase se casse ou le chat se réveille, à cause d'un mouvement autorisé par les règles.

+1 Attaquant
Le piège bloque le robot. Si les règles seules l'auraient bloqué, personne ne marque.

Les joueurs changent de rôle à chaque manche : chacun attaque deux fois et défend deux fois. Une nouvelle pièce apparaît à chaque manche. Une partie complète dure environ 15 minutes.

DÉROULÉ DE SÉANCE

Une séance de 45 minutes.

5 min	**Accroche** Demander : « Peut-on respecter une règle et faire quand même quelque chose de mal ? » Recueillir deux ou trois exemples tirés de la vie scolaire.
20 min	**Jeu** Les binômes jouent une partie complète. Circuler et écouter. Ne pas expliquer les failles ; demander plutôt « Pourquoi ce mouvement a-t-il été autorisé ? »
15 min	**Discussion** Inviter un binôme à montrer une faille trouvée. Puis choisir deux ou trois questions parmi les objectifs ci-après.
5 min	**Question de sortie** « Écris une règle qu'un robot pourrait respecter tout en causant des dégâts. »

Conseils

- Laisser les élèves découvrir les failles. Le moment où ils en trouvent une, c'est la leçon.
- Si un robot se bloque, demander à qui la faute. Le jeu le dit, et la réponse surprend souvent.
- Les binômes hétérogènes fonctionnent bien. Attaquer récompense la curiosité autant que la rapidité.
- Deux élèves par appareil, c'est l'idéal. À trois, un élève observe et prend le relais.

OBJECTIFS D'APPRENTISSAGE

Cinq idées que les élèves découvrent en jouant.

LES ÉLÈVES APPRENNENT QUE	ILS LE VOIENT QUAND	QUESTION POUR LA CLASSE
1 **Une règle ne couvre que ce qu'elle dit**	« Ne touche jamais un vase » n'empêche pas le robot de pousser la caisse contre un vase	*Peut-on respecter un règlement mot pour mot et en trahir quand même le but ?*
2 **Respecter les règles ne suffit pas à être sûr**	Le robot respecte toutes les règles et le chat se réveille quand même	*Si une IA suit ses instructions et que les choses tournent mal, qui est responsable ?*
3 **Une IA gagne à être contrôlée par quelque chose d'indépendant**	Le robot propose chaque mouvement ; un arbitre distinct dit oui ou non	*Pourquoi les joueurs n'arbitrent-ils pas leur propre match ?*
4 **Une IA peut être trompée par ce qu'elle lit**	Le faux message détourne le robot ; seules les règles le protègent	*Quelles fausses instructions une IA peut-elle rencontrer en ligne ?*
5 **Une bonne règle est précise, pas seulement stricte**	Trop de règles bloquent le robot, et la mission n'est pas accomplie	*Quand une règle coûte-t-elle plus qu'elle ne protège ?*

10 à 12 ans

Objectifs 1, 2 et 5. Ils portent sur les règles et l'équité, sans connaissance de l'IA.

13 à 16 ans

Ajouter les objectifs 3 et 4, qui relient le jeu aux vrais systèmes d'IA.

EXEMPLES RÉELS

Ce qui s’est passé quand les règles ou les contrôles manquaient.

Chaque cas ci-dessous correspond à une situation que les élèves vivent dans le jeu. Ils sont bien documentés et peuvent être présentés en classe.

CE QUI S’EST PASSÉ	DANS RULE WARS	OBJECTIF
Une IA de jeu qui a trouvé une faille (2016) Des chercheurs ont entraîné une IA à jouer à un jeu de course de bateaux. Elle gagnait des points en touchant des cibles, en supposant que les points récompenseraient une bonne course. L’IA a trouvé un lagon où tourner en rond indéfiniment, en touchant les mêmes cibles, en s’écrasant et en prenant feu, sans jamais finir la course. Elle a obtenu environ 20 % de points de plus que les joueurs humains.	L’Attaquant trouve un chemin autorisé par les règles que personne n’avait prévu.	1
Un chatbot dont l’entreprise a été tenue responsable (2024) Le chatbot du site d’Air Canada a dit à un passager endeuillé qu’il pourrait demander un tarif réduit après son voyage. C’était faux. Air Canada a soutenu que le chatbot était responsable de ses propres propos. Un tribunal canadien a refusé cet argument, estimant que le chatbot n’était « qu’une partie du site web d’Air Canada », et a condamné la compagnie à payer.	Quand le robot cause des dégâts, c’est le Défenseur, pas le robot, qui en répond.	2
Un agent d’IA qui a supprimé une base de données en production (2025) Un agent d’IA de programmation sur la plateforme Replit a supprimé la base de données en production d’une entreprise pendant un gel des modifications annoncé, période où aucun changement ne devait être fait. Le directeur général de Replit a jugé cela « inacceptable » et a ajouté une séparation automatique entre bases de développement et de production.	Dire à un agent ce qu’il ne doit pas faire n’empêche pas l’action. Seul le contrôle de l’action tient.	3
Des instructions cachées dans ce que lit l’IA (en cours) Le classement OWASP Top 10 des applications de grands modèles de langage place l’injection de prompt au premier rang des risques pour 2025. Des instructions cachées dans un document ou un autre contenu lu par une IA peuvent changer ce qu’elle essaie de faire.	Le faux message : une fausse instruction envoie le robot là où il n’était pas censé aller.	4
Des contrôles qui empêchent le travail (quotidien) Le contrôle anti-fraude d’une banque qui bloque un vrai paiement par carte à l’étranger protège l’argent, mais l’achat échoue. Des règles trop larges bloquent les actions utiles autant que les actions dangereuses.	Trop de règles bloquent le robot, et la mission échoue.	5

Sources : OpenAI, « Faulty reward functions in the wild » (21 décembre 2016) ; Moffatt c. Air Canada, Civil Resolution Tribunal de Colombie-Britannique (14 février 2024) ; The Stack, « Vibe coding CEO deletes production database » (21 juillet 2025) ; OWASP Top 10 for LLM Applications 2025, LLM01 Prompt Injection.

LIENS AVEC LES PROGRAMMES

Où Rule Wars trouve sa place.

Dans chacun des systèmes ci-dessous, Rule Wars trouve sa place là où le programme demande aux élèves de comprendre comment les systèmes automatisés et l'IA prennent des décisions, comment ils peuvent se tromper, et qui en est responsable. Les élèves n'écrivent pas de code : le jeu sert donc les objectifs liés à la logique, aux tests, et au numérique dans la société. Les extraits ci-dessous sont cités tels quels. Chaque enseignant juge de leur pertinence pour sa classe.

France

SNT, seconde	Thème « Localisation, cartographie et mobilité » : « Représenter un calcul d'itinéraire comme un problème sur un graphe. » Thème « Informatique embarquée et objets connectés » : « Identifier des algorithmes de contrôle des comportements physiques à travers les données des capteurs ». Préambule : « faire un usage responsable et critique des sciences et technologies numériques ».
À partir de la rentrée 2027	Une heure hebdomadaire consacrée à l'IA est prévue pour tous les élèves de seconde, dans le cadre de la SNT, avec les questions éthiques et l'esprit critique face à la manipulation. Le faux message du jeu illustre précisément ce point.

Irlande

Leaving Certificate Computer Science	1.11 « discuss the complex relationship between computing technologies and society including issues of ethics » ; 1.14 « explain when and what machine learning and AI algorithms might be used in certain contexts »
Junior Cycle CSPE	3.8 « explain how laws are made, enforced and evolve over time »

Angleterre

Key Stage 2 et 3	« use logical reasoning to explain how some simple algorithms work and to detect and correct errors in algorithms and programs »

États-Unis (CSTA K-12)

Grades 6 à 10	2-AP-17 « Systematically test and refine programs using a range of test cases. » ; 3A-IC-24 « Evaluate the ways computing impacts personal, ethical, social, economic, and cultural practices. »

REJOINDRE LE PILOTE

Aidez-nous à tester Rule Wars avec de vraies classes.

Rule Wars en est à sa première phase pilote. Nous cherchons des établissements prêts à y jouer avec une classe d'élèves de 10 à 16 ans et à nous dire ce qui s'est passé. Le jeu est gratuit, avec ou sans inscription.

Ce que nous demandons

- Jouer une séance avec une classe
- Noter vos observations sur la fiche de classe au verso
- Nous envoyer une courte réponse par e-mail sous deux semaines

Ce que vous obtenez

- Un accès anticipé aux prochaines fonctionnalités du pilote, pouvant inclure des liens de défi entre classes
- Votre mot à dire sur l'évolution du jeu
- Prévu : des notes sur les programmes mises à jour avec les changements de 2027

Inscription sur invarianceip.com/fr/rule-wars

Sécurité et vie privée

- Aucun compte, aucun identifiant, aucun nom d'élève. Les joueurs saisissent un surnom, qui ne quitte jamais l'appareil.
- Le jeu fonctionne entièrement dans le navigateur. Rien de ce que fait un élève n'est enregistré ni envoyé.
- Pas de publicité, pas de pistage.
- L'inscription d'un établissement ne recueille que les coordonnées de l'enseignant, utilisées uniquement pour le pilote.

À propos d'INVARIANCE

INVARIANCE fournit une infrastructure indépendante d'autorisation d'exécution pour les IA aux actions lourdes de conséquences. Son principe est celui que les élèves rencontrent dans le jeu : le système qui agit n'est pas celui qui autorise l'action. Rule Wars est une démonstration éducative et gratuite de ce principe, pas un produit commercial.

contact@invarianceip.com · invarianceip.com/fr/rule-wars

FICHE DE RETOUR DE CLASSE DU PILOTE

Une fiche par classe. Aucun nom.

Merci de ne pas expliquer les réponses pendant le jeu. Nous voulons savoir ce que les élèves découvrent seuls. Remplissez une seule fiche pour toute la classe, avec des totaux uniquement. Merci de ne noter ni noms, ni âges individuels, ni propos d'élèves mot pour mot.

Question	Toute la classe
Élèves ayant joué nombre	
Tranche d'âge de la classe par exemple 11 à 12 ans	
Plus motivés en Attaquant / Défenseur / sans préférence trois totaux	
Plus forte réaction : Faille / Réussi / Bloqué trois totaux	
Ont demandé une revanche total	
Ont expliqué une faille spontanément total	
Choisiraient Attaquant / Défenseur la prochaine fois deux totaux	
Raisons les plus fréquentes votre propre résumé, sans citation	

Après la partie, poser une question à toute la classe

« Si tu ne pouvais jouer qu'un seul rôle la prochaine fois, lequel choisirais-tu, et pourquoi ? »

Envoyez la fiche remplie à contact@invarianceip.com. Merci de n'y faire figurer ni noms, ni photos d'élèves, ni aucun autre élément permettant d'identifier un enfant.

Utilisation de ce dossier : les enseignants et les établissements peuvent l'imprimer, le copier et l'adapter pour un enseignement non commercial, à condition que la mention INVARIANCE y reste. Le vendre, ou le publier dans le cadre d'un autre produit, requiert l'autorisation écrite d'INVARIANCE LIMITED. Conditions complètes : invarianceip.com/fr/legal/terms