Eigenflow

GUIDE

MiroFish, la foule simulée, refaite avec Claude

Ce que fait le simulateur, ce qu'il coûte, où partent tes données, notre test avec 100 personnes et deux IA, et les cinq prompts pour tester ton idée.

PAR

Abyane Nabeebaccus

INSTAGRAM

@abyane_nb

ÉDITION

novembre 2026

Tu as commenté FOULE, voici le guide. MiroFish est un logiciel libre qui fabrique une petite société d'agents IA à partir d'un texte, les fait parler sur deux faux réseaux sociaux, puis écrit un rapport sur ce qui s'est passé. Le dépôt, github.com/666ghj/MiroFish, est sous licence AGPL-3.0 et comptait 84 883 étoiles le 8 novembre 2026 au soir.

D'abord ce que fait MiroFish et ce qu'il coûte, données comprises. Ensuite notre test : son principe refait avec Claude seul, 100 personnes fictives sur deux tours de réseau social, joué par deux modèles qui ne sont pas d'accord. Les cinq prompts sont donnés mot pour mot, avec des crochets pour ta propre idée.

Dépôt, FAQ officielle, presse et articles relus le 8 novembre 2026. Eigenflow n'a aucun lien avec MiroFish, Shanda, Alibaba Cloud, Zep ni Anthropic.

À savoir avant de lire

- On n'a pas installé MiroFish : c'est du code tiers, et il demande deux comptes en ligne payés à l'usage. Ce qui le concerne vient du README, du code de la branche main et des tickets du dépôt. Notre test refait son principe avec Claude.
- Le README promet des milliers d'agents. Ses propres démos en ont 39, 55 et 68, et un test extérieur en a obtenu 25 en en voulant 100.
- Il faut une clé d'API d'IA payée à l'usage et un compte Zep Cloud. L'appli annonce 5 $ en moyenne par simulation. Un abonnement Claude ou ChatGPT ne peut pas le faire tourner.
- Tout ce que tu déposes part chez le fournisseur d'IA et chez Zep Cloud. Pas de données de clients, pas de document confidentiel.
- Aucune étude ne montre qu'il prédit juste. Sur le seul banc d'essai indépendant trouvé, il obtient 63,1 sur 100, et l'IA seule 63,4.

1. Ce qu'est MiroFish

Qui l'a fait

L'auteur s'appelle Guo Hangjiang, BaiFu sur GitHub (compte 666ghj). Le Science and Technology Daily, quotidien d'État chinois, le présente le 9 mars 2026 comme « 中国大四学生 », un étudiant chinois de 4e année, la dernière de la licence, qui a développé MiroFish seul. D'après ChinaVenture (13 mars 2026), il étudie à l'Université des postes et télécommunications de Pékin. Le Xinmin Evening News (11 mars 2026) précise qu'il l'a construit pendant un stage chez Shanda. Sur GitHub, il signe 266 des 322 commits ; un second contributeur, Ghostubborn, en signe 34.

Le dépôt est créé le 26 novembre 2025. La v0.1.2, dernière version publiée, sort le 7 mars 2026, et le projet arrive le même jour en tête de GitHub Trending, toutes langues confondues : l'archive de la Wayback Machine de 19 h 23 (UTC) le montre premier avec « 345 stars today ». Il y revient le 16 mars avec 3 257 étoiles dans la journée. Le record tombe le 9 mars : 3 703 étoiles gagnées en un jour, d'après la courbe que le dépôt publie lui-même.

L'argent de Shanda

Shanda est le groupe de Chen Tianqiao, devenu milliardaire avec les jeux en ligne, Legend of Mir en tête. Le classement Hurun du 27 octobre 2026 place le couple qu'il forme avec Luo Qianqian à 10,7 milliards de dollars. Le README dit seulement que MiroFish « has received strategic support and incubation from Shanda Group », sans montant. Le chiffre de 30 millions de yuans vient de la presse : Science and Technology Daily (9 mars 2026) et Xinmin Evening News (11 mars 2026). Au cours de la BCE du 6 novembre 2026 (7,4485 yuans pour 1 €), ça fait environ 4 millions d'euros. Aucune annonce officielle de Shanda ne donne ce montant, et le Wenhui Daily parle, lui, d'une valorisation de 30 millions.

Ce que fait chaque étape

Le README découpe le travail en cinq étapes. Voici ce que fait le code de la branche main, lu sans l'exécuter.

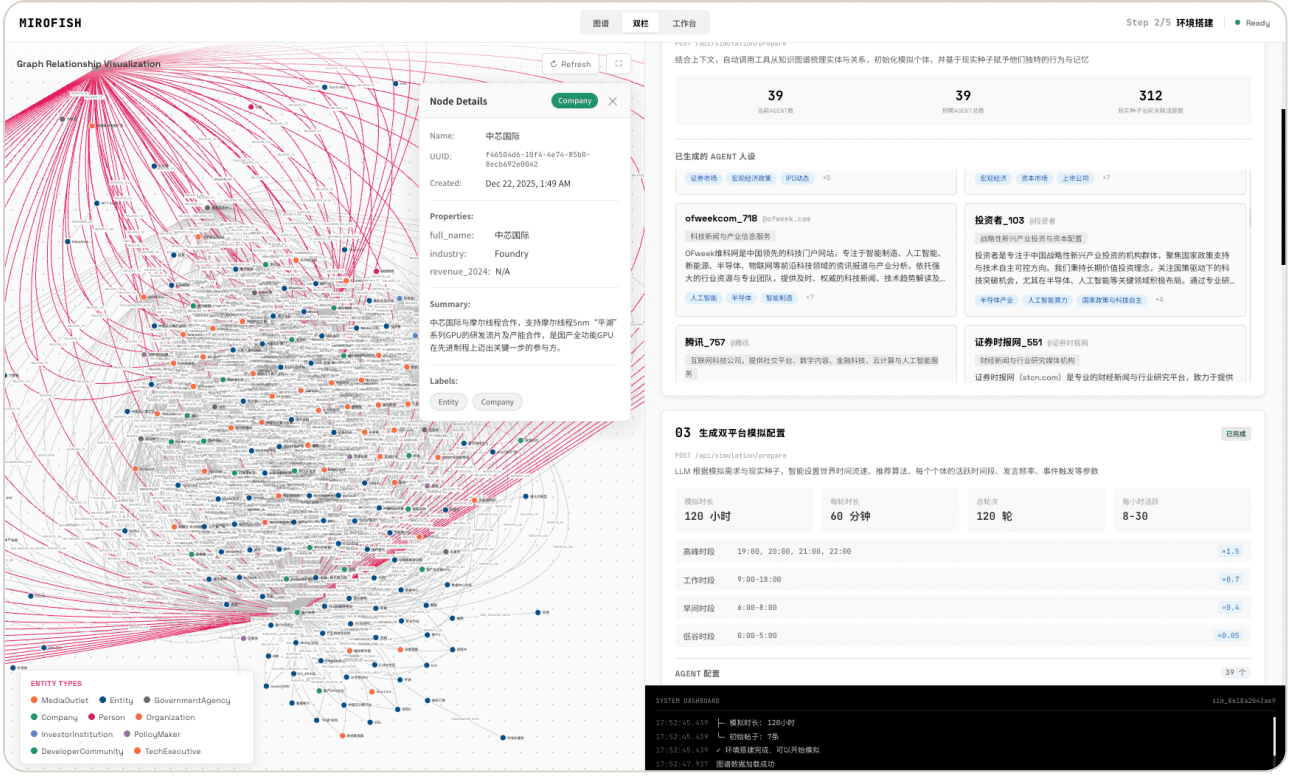
1. Le graphe. Tu déposes un fichier et tu écris ta question. L'IA définit exactement 10 types d'entités (personne, entreprise, média, administration…), puis Zep Cloud, un service en ligne, construit un graphe de connaissances à partir de ton texte.

2. Les fiches. Chaque entité du graphe devient un agent. Pour une personne, l'IA écrit une fiche d'environ 2 000 caractères : âge, métier, profil MBTI, positions, tics de langage, souvenirs de l'événement. Pour une organisation, l'âge est fixé à 30 ans.

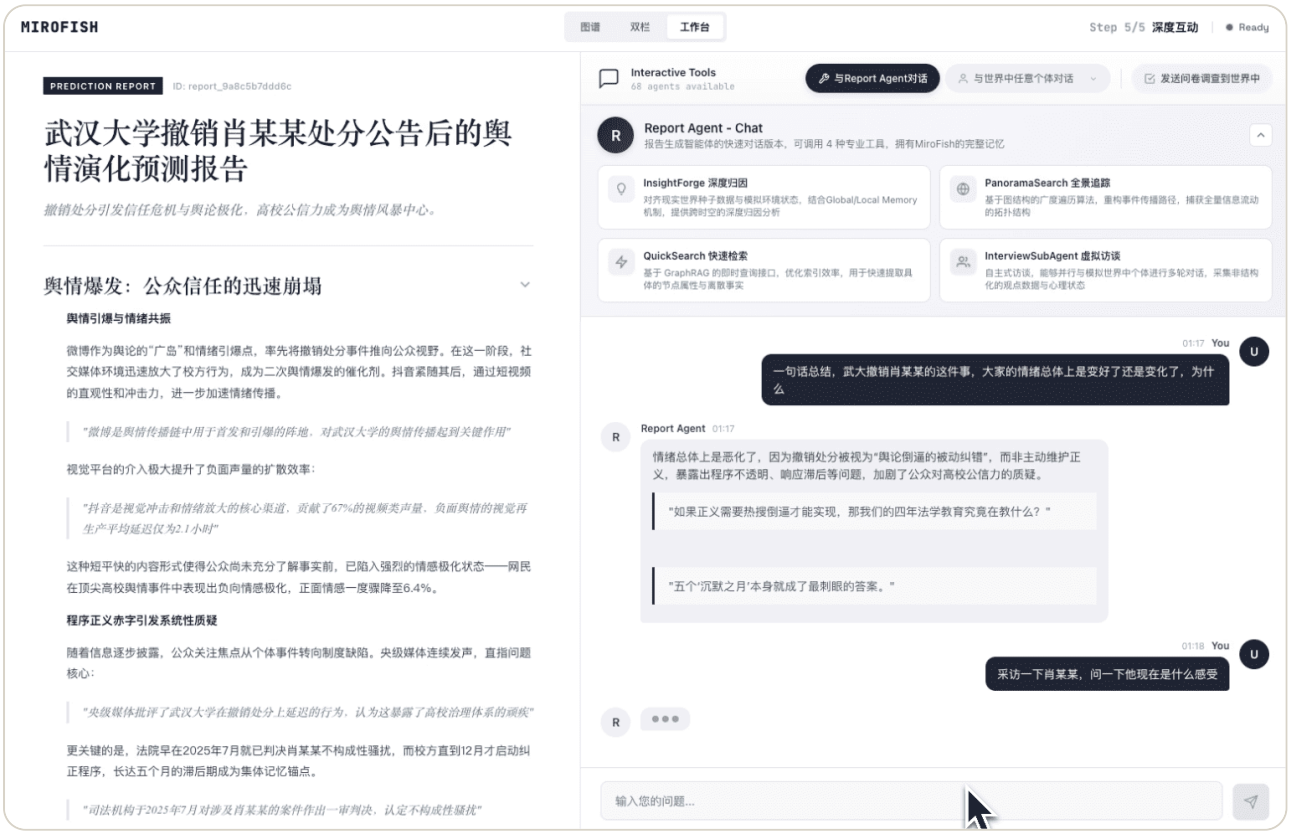
3. La simulation. Deux réseaux tournent en parallèle grâce à OASIS, un moteur libre de l'équipe CAMEL-AI : un façon Twitter (poster, aimer, reposter, citer, suivre) et un façon Reddit (poster, commenter, voter, chercher, suivre, masquer). Un tour vaut une heure simulée, 10 tours par défaut.

4. Le rapport. Un agent appelé ReportAgent écrit de 2 à 5 chapitres. Il fouille la simulation avec quatre outils et interroge lui-même jusqu'à 5 agents.

5. Les entretiens. Tu peux parler à n'importe quel agent, ou envoyer un questionnaire à plusieurs d'un coup, tant que la simulation tourne encore (FAQ officielle, question 20).



Étape 2 de MiroFish, capture du README (licence AGPL-3.0) : à gauche le graphe tiré d'un texte sur la levée de fonds de Moore Threads, à droite les fiches d'agents générées. Les compteurs indiquent 39 agents créés sur 39 prévus.



Étape 5, capture du README : le rapport de la démo sur l'université de Wuhan, le chat avec le ReportAgent, « 68 agents available » en haut, et le bouton « 与世界中任意个体对话 » (parler à n'importe quel individu du monde).

---

2. Ce que la vidéo simplifie

Une vidéo courte coupe des détails. Voici ceux qui comptent avant d'installer quoi que ce soit.

- Des dizaines d'agents. Le README annonce « thousands of intelligent agents », la page d'accueil de l'appli « jusqu'à un million ». Les exemples officiels tournent avec 39 agents

(capture 2 du README), 55 (la démo en ligne figée sur Wuhan) et 68 (capture 5). WPP Research, qui l'a testé en mai 2026, écrit : « We wanted 100 agents; we got around 25. » Aucun réglage ne permet de choisir ce nombre, il dépend des entités que l'IA trouve dans ton texte.

- Il faut un fichier. Le bouton de lancement exige un texte de demande et au moins un fichier PDF, MD ou TXT. Tu peux écrire ton idée dans un .txt, mais la formulation pèse lourd : dans le ticket 838 (3 octobre 2026), la même question donne 2 agents (« GitHub » et « Zep Cloud ») ou 8 selon la façon d'écrire le texte de départ.
- Les agents sont les acteurs nommés dans ton texte. La consigne envoyée à l'IA exige des acteurs « qui existent réellement et peuvent s'exprimer sur les réseaux sociaux ». Entreprises et médias deviennent des agents : dans la démo officielle, le premier de la liste s'appelle « Facebook ».
- « Des amis » est un compteur. Les champs friend_count et follower_count sont tirés au hasard (entre 50 et 500, et entre 100 et 1 000). Les vrais liens viennent du graphe et de l'action « suivre » pendant la simulation.
- « Changent d'avis » n'est mesuré nulle part. Aucune variable d'opinion n'est mise à jour pendant la simulation. Les agents lisent les posts des autres, un revirement peut apparaître dans leurs messages, mais rien ne le compte.
- Zep Cloud est obligatoire. Le logiciel refuse de démarrer sans clé Zep, et aucune alternative locale n'est intégrée (FAQ, question 11).
- Ça coûte. L'appli affiche « 低成本 · 常规模拟平均5$/次 » : bas coût, 5 $ par simulation en moyenne. Le README prévient : « High consumption, try simulations with fewer than 40 rounds first ».
- Le modèle recommandé est qwen-plus d'Alibaba Cloud. Les autres modèles compatibles OpenAI ne sont pas tous garantis (FAQ, question 8).



Accueil de MiroFish, capture du README : « 低成本 · 常规模拟平均5$/次 » (bas coût, 5 $ par simulation en moyenne) et « 最多百万级Agent模拟 » (jusqu'à un million d'agents simulés). À droite, la zone de dépôt n'accepte que des fichiers PDF, MD ou TXT.

3. Installer MiroFish (non testé par nous)

Ces étapes recopient le README et la FAQ, relus le 8 novembre 2026. On ne les a pas exécutées.

Ce qu'il te faut

- Node.js 18 ou plus, Python 3.11 ou 3.12 (la FAQ déconseille Python 3.13) et uv, le gestionnaire de paquets Python.
- Une clé d'API d'un modèle compatible avec le format OpenAI. La FAQ (question 9) est claire : un abonnement ChatGPT, une connexion Codex ou une autre formule de chat ne marche pas, il faut une vraie clé d'API. Le branchement direct sur l'API d'Anthropic n'est pas dans la branche main (question 8).
- Une clé Zep Cloud, à créer sur app.getzep.com.

Les commandes

La FAQ (question 3) précise que la branche main a 96 commits d'avance sur la v0.1.2, dont des correctifs de sécurité. Le README recommande donc l'installation depuis les sources.

> git clone https://github.com/666ghj/MiroFish.git

> cd MiroFish

> cp .env.example .env

Ouvre ensuite le fichier .env et remplis les quatre lignes. Celles du README visent dashscope.aliyuncs.com, l'adresse de la région Chine (Pékin) d'Alibaba Cloud, là où sont stockées les données des requêtes d'après la documentation d'Alibaba (mise à jour le 6 novembre 2026). La région Singapour répond sur dashscope-intl.aliyuncs.com, avec une clé créée dans cette région :

```
À COLLER

LLM_API_KEY=ta_cle_alibaba
LLM_BASE_URL=https://dashscope-intl.aliyuncs.com/compatible-mode/v1
LLM_MODEL_NAME=qwen-plus
ZEP_API_KEY=ta_cle_zep
```

> npm run setup:all

> npm run dev

L'interface s'ouvre sur http://localhost:3000, l'API sur le port 5001. Avec Docker, après le même fichier .env, c'est `docker compose up -d`.

Ce que ça coûte

Prix relevés le 8 novembre 2026, affichés en dollars, convertis au cours BCE du 6 novembre (1 € pour 1,1165 $), avant une éventuelle TVA :

- L'appli annonce 5 $ par simulation en moyenne, environ 4,50 €. La FAQ (question 22) ajoute qu'il n'existe pas de chiffre fixe et que seul le tableau de bord de ton fournisseur fait foi.

- qwen-plus, tarif international d'Alibaba Cloud (page mise à jour le 7 novembre 2026) : 0,40 $ le million de jetons en entrée et 1,20 $ en sortie, soit 0,36 € et 1,07 €.
- Zep Cloud : 10 000 crédits gratuits par mois, sans report d'un mois sur l'autre, puis la formule Flex à 125 $ par mois (environ 112 €). Un crédit couvre 350 octets de texte envoyé. WPP Research a vidé le plan gratuit en trois lancements, à environ 1 $ par lancement avec gemini-2.5-flash-lite et 25 agents.

Méfie-toi des sites qui vendent des abonnements MiroFish. La FAQ (question 2) cite mirofish.my, mirofish.homes et mirofish.co.in, qui n'ont rien d'officiel. Le site mirofish.ai n'affiche qu'une liste d'attente, et la démo du README est figée : elle n'est branchée à aucune IA.

Tes données : RGPD et sécurité

Le logiciel tourne chez toi, mais ton texte en sort. Il part chez le fournisseur du modèle, Alibaba Cloud si tu suis le README, et chez Zep AI, Inc., qui construit le graphe et garde la mémoire des agents pendant la simulation. La FAQ de la page de prix de Zep (8 novembre 2026) dit qu'elle signe un accord de traitement des données avec ses clients européens, mais son tableau comparatif ne le coche que pour la formule Enterprise : demande-le avant d'envoyer quoi que ce soit.

La CNIL, dans ses questions-réponses sur l'IA générative du 18 juillet 2024, conseille d'« interdire la fourniture de certaines données confidentielles (par exemple couvertes par le secret industriel et commercial) ou personnelles ». Elle rappelle aussi que si des données personnelles peuvent partir hors de l'Union européenne, « l'entité utilisatrice devra, en complément, encadrer leur traitement par leur destinataire ». Côté texte, c'est l'article 28 du RGPD pour le contrat avec un sous-traitant et le chapitre V (articles 44 à 49) pour les transferts hors de l'Union.

Notre règle : ni nom, ni e-mail, ni historique d'achat de vrais clients dans le fichier. Pas de document interne confidentiel non plus. Décris ton idée et ta cible avec des mots génériques. Et puisque MiroFish transforme en agents les acteurs nommés dans ton texte, n'y écris pas le nom de vraies personnes : il en ferait des personnages.

Côté sécurité, la faille CVE-2026-7042, publiée par le NVD le 26 avril 2026, signale une API sans authentification jusqu'à la v0.1.2. La FAQ (question 16) confirme que la branche main écoute par défaut sur toutes les adresses réseau, accepte les requêtes de n'importe quel site et n'a aucune gestion d'utilisateurs : « Use the default project only on a trusted personal/local network. » Lance-le chez toi, jamais sur le wifi d'un café ou d'un espace de coworking, et arrête-le après usage.

Dernier point, la licence. L'AGPL-3.0 te laisse utiliser MiroFish librement. Si tu modifies le code et que d'autres l'utilisent par le réseau (un service en ligne pour tes clients, par exemple), tu dois leur proposer le code source de ta version (article 13 de la licence).

---

4. Notre test : MiroFish refait avec Claude seul

On a refait le principe de MiroFish avec Claude, en ligne de commande (claude -p, abonnement Claude, aucun outil, dans un dossier vide pour qu'aucun réglage du projet ne soit lu). Le 8 novembre 2026, avec cette idée :

« Une appli de repas healthy par abonnement : chaque semaine tu choisis 5 déjeuners équilibrés, on te les livre chez toi ou au bureau. 44,90 € la semaine, livraison 3,90 €. »

La méthode

1. Claude Sonnet 5 génère 100 personnes fictives représentatives de la population française adulte, en 4 lots de 25, une seule fois. Chacune a un prénom, un âge, un métier, une ville, une situation, trois mots de caractère et ses habitudes de repas.

2. Le code, sans IA, relie chaque personne à ses amis : ses deux voisins de chaque côté dans la liste, plus un lien tiré au hasard pour un tiers des personnes environ (graine fixe). Chacun a de 4 à 7 amis, 4,6 en moyenne.

3. Tour 1 : chacun découvre l'idée en post sponsorisé et poste, partage ou ignore, avec un score de -2 (rejette franchement) à +2 (adore, prête à payer).

4. Tour 2 : chacun relit son premier avis et les posts de ses amis, commente sous le post d'un ami et redonne un score.

5. Un agent écrit le rapport : qui a aimé, qui a rejeté, où ça casse.

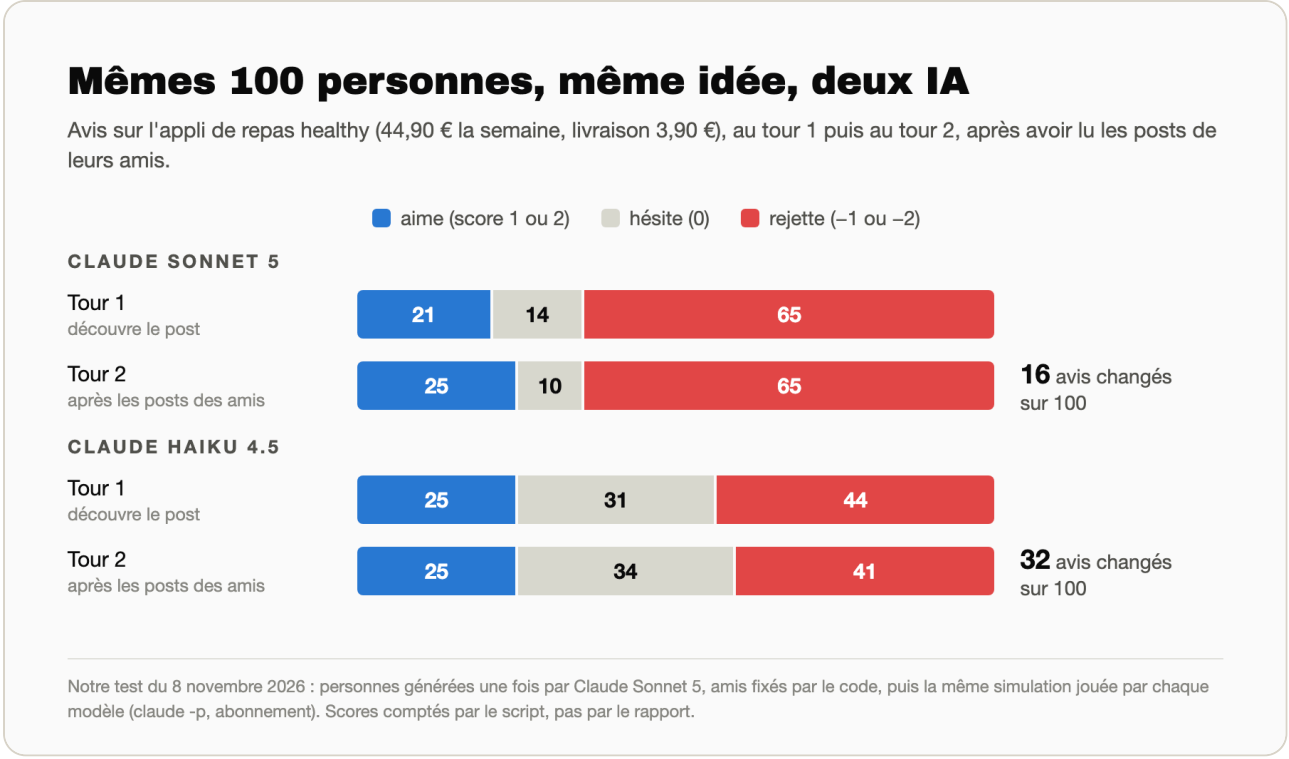
6. Entretien : on demande à une personne qui rejette « Qu'est-ce qui te ferait changer d'avis et t'abonner ? ».

7. On rejoue les étapes 3 à 6 avec Claude Haiku 4.5, sur les mêmes personnes et les mêmes amis.

Les scores sont comptés par le script, jamais par le rapport. On a appris pourquoi en lisant celui de Haiku (plus bas).

Ce qu'on a obtenu

Avec Sonnet, 52 personnes postent au premier tour, 7 partagent et 41 ignorent. Avec Haiku, 77 ignorent, 15 partagent et 8 postent. Les avis :



Notre test du 8 novembre 2026 : avis des 100 personnes au tour 1 puis au tour 2. Sonnet : 21 aiment, 14 hésitent, 65 rejettent, puis 25, 10 et 65, avec 16 avis changés. Haiku : 25, 31 et 44, puis 25, 34 et 41, avec 32 avis changés.

Le second tour bouge peu chez Sonnet : 16 avis changent, 10 vers le haut et 6 vers le bas. Le rapport retient le cas de Camille, 24 ans, vendeuse à Lille, qui passe de -1 à +1 sous le post de Delphine, avocate à Paris : « Enfin je vais arrêter de commander n'importe quoi à 23h après une audience, j'essaie dès cette semaine. » Chez Haiku, 32 avis changent, deux fois plus.

Entre les deux modèles, 64 personnes sur 100 finissent dans le même camp au second tour (40 rejettent chez les deux, 18 aiment chez les deux, 6 hésitent chez les deux). Les 36 autres changent de camp quand on change d'IA, et quatre passent carrément de « rejette » à « aime » ou l'inverse. C'est le résultat qui nous a le plus surpris. Nadia en fait partie.



Nadia, 35 ans, infirmière à Marseille : même fiche, mêmes amis. Jouée par Sonnet, elle rejette (-1) ; jouée par Haiku, elle adore (+2). Textes bruts du test, fautes comprises.

À l'entretien de la version Sonnet, Nadia répond en 6 secondes : « Franchement ? Si c'était 15-20 euros la semaine, livraison gratuite, et adapté aux enfants aussi. Sinon je préfère garder cet argent

pour autre chose, mes tupperwares font le job. »

Chez Haiku, Nadia adore l'idée, alors le script a interrogé le premier parent qui rejette : Bernard, 58 ans, plombier à Clermont-Ferrand. Sa condition : « faudrait réduire les prix de moitié, des portions qui remplissent vraiment un homme qui bosse, et que ça goûte bon ».

Ce que la foule a dit du prix

Les deux rapports désignent le prix. Sonnet écrit « Prix (44,90€+livraison) perçu comme inaccessible face aux salaires modestes », Haiku « Budget : cible riche, exclut 50% de la population précaire ». Ce « 50 % », Haiku l'a inventé : rien dans la simulation ne le calcule. C'est la raison pour laquelle on compte les scores avec du code.

Au premier tour de Sonnet, 36 des 65 rejets donnent l'argent comme raison (cher, budget, salaire, bourse…, compté par mots-clés dans le champ « raison »). Dix posts parlent de « 48 € » : les personnes ont additionné la livraison toutes seules (44,90 + 3,90 = 48,80 €). Les autres rejets parlent de gamelles maison, de potager, d'horaires en 3x8 ou de déjeuners pris dans un camion ou sur un chantier, là où une livraison « au bureau » ne sert à rien.

Le rapport de Sonnet range ceux qui aiment en trois groupes : « avocats et cadres surbookés », « infirmières et soignantes libérales », « jeunes actifs urbains aisés ». Et ceux qui rejettent : « retraités attachés au fait-maison », « foyers modestes et familles nombreuses », « jeunes précaires, étudiants, apprentis ».

Notre avis : la foule a trouvé l'objection du prix, que n'importe quel commerçant aurait vue en dix secondes. Ce qu'elle apporte en plus, ce sont les mots des gens (« 48 € », « mes tupperwares ») et des objections plus rares, comme le déjeuner pris dans un camion.

Un défaut à surveiller : on avait demandé des adultes de 18 à 75 ans, et Claude a glissé trois lycéens de 16 et 17 ans. Relis toujours le fichier des personnes avant de lancer les tours.

Ce que ça a coûté

Tout le test (100 personnes, deux simulations complètes, 24 appels) a tourné sur l'abonnement Claude, sans frais en plus. Au tarif de l'API, Claude Code l'estime à 1,69 $ (environ 1,51 € hors taxes) : 0,17 $ pour les personnes, 0,81 $ pour la simulation Sonnet, 0,71 $ pour celle de Haiku. Avec 4 lots en parallèle, la simulation Sonnet a pris moins de 3 minutes, celle de Haiku un peu plus de 5.

---

5. Refais-le avec Claude : les cinq prompts

Ce sont nos prompts du 8 novembre, mot pour mot, avec des crochets à remplacer. Deux façons de t'en servir.

- Dans claude.ai, avec 25 personnes : tu colles les prompts un par un. Ouvre une conversation neuve à chaque étape et recolle les tableaux JSON de l'étape d'avant. C'est ce que fait notre script : chaque appel part de zéro, sans la mémoire du précédent.
- Dans Claude Code, avec 100 personnes : un script appelle Claude à chaque étape et compte les résultats (prompt plus bas).

Garde chaque réponse dans un fichier texte (personnes.json, tour1.json, tour2.json). Ce sont eux que tu recolles.

Prompt 1, les personnes

Remplace la cible et le dernier champ par ce qui touche ton produit. Pour notre appli de repas, c'était « habitudes_repas ».

> Tu génères des personnes fictives pour une simulation de réaction du public en France. Génère 25 personnes très différentes les unes des autres, numérotées de 1 à 25, représentatives de [TA CIBLE, par exemple : la population française adulte (âges de 18 à 75 ans, villes et campagnes, revenus modestes à aisés, métiers variés, situations familiales variées)]. Elles ne savent rien d'aucun produit. Réponds uniquement par un tableau JSON, sans texte autour, chaque objet : {"id": nombre, "prenom": "...", "age": nombre, "metier": "...", "ville": "...", "situation": "une phrase : famille, budget, rythme de vie", "caractere": "trois mots", "habitudes": "une phrase sur [CE QUI TOUCHE TON PRODUIT, par exemple ses déjeuners en semaine]"}.

Pour aller jusqu'à 100, relance-le trois fois en changeant la numérotation (26 à 50, et ainsi de suite) et ajoute à la fin « Lot 2 sur 4 : évite les doublons de profils évidents. ».

Prompt 2, le premier tour

> Simulation d'un réseau social. Tu joues, une par une et séparément, chacune des personnes du tableau JSON ci-dessous. Chacune découvre ce post sponsorisé dans son fil : « [TON IDÉE, AVEC SON PRIX] ». Pour chaque personne, décide ce qu'elle fait vraiment, avec sa voix à elle (registre de réseau social, tutoiement possible, fautes et humeur permises, pas de langue de bois). Elle peut ignorer. Réponds uniquement par un tableau JSON, un objet par personne : {"id": nombre, "action": "poste" | "partage" | "ignore", "texte": "son post en 25 mots maximum, vide si ignore", "score": entier de -2 (rejette franchement) à 2 (adore, prête à payer), "raison": "5 mots maximum"}. Personnes : [COLLE ICI LE TABLEAU JSON DES PERSONNES]

Prompt 3, le second tour

Dans claude.ai, les amis suivent une règle simple que tout le monde peut recalculer : les deux numéros d'avant et les deux d'après. Notre script ajoutait en plus quelques liens au hasard.

> Suite de la simulation d'un réseau social. Le produit dont tout le monde parle : « [TON IDÉE, AVEC SON PRIX] ». Tu joues, une par une et séparément, chacune des personnes du premier tableau JSON ci-dessous. Pour chacune : ce qu'elle avait pensé au premier tour (second tableau), puis ce que ses amis ont publié. Les amis de la personne n sont les personnes n-2, n-1, n+1 et n+2 (après la dernière on repart à 1, avant la première on repart à la dernière) ; un ami qui a ignoré le post n'a rien publié. Elle réagit sous le post d'UN ami (celui qui la fait réagir le plus) et peut changer d'avis ou se braquer, comme dans la vraie vie. Si aucun de ses amis n'a publié, "repond_a" vaut null et elle donne seulement son avis. Réponds uniquement par un tableau JSON, un objet par personne : {"id": nombre, "repond_a": id de l'ami, "commentaire": "20 mots maximum, sa voix", "score": entier de -2 à 2 (son avis maintenant), "pourquoi": "si son avis a bougé, 8 mots maximum, sinon vide"}. Personnes : [TABLEAU DES PERSONNES] Premier tour : [TABLEAU DU TOUR 1]

Prompt 4, le rapport

> Tu es l'agent qui écrit le rapport final d'une simulation. [NOMBRE] personnes fictives ont découvert ce produit sur un réseau social : « [TON IDÉE, AVEC SON PRIX] ». Voici qui elles sont, puis tout ce qui s'est passé au premier et au second tour (score de -2 à 2) : [TABLEAU DES PERSONNES] [TABLEAU DU TOUR 1] [TABLEAU DU TOUR 2]. Écris le rapport pour le créateur du produit. Réponds uniquement par un objet JSON : {"verdict": "une phrase", "ont_aime": [{"qui": "groupe en 4 mots", "pourquoi": "8 mots max"}] (3 max), "ont_rejete": [{"qui": "groupe en 4 mots", "pourquoi": "8 mots max"}] (3 max), "ou_ca_casse": "la principale raison de rejet, 10 mots max", "changement_notable": "un changement d'avis marquant, 15 mots max"}

Compte les scores toi-même : 1 et 2 aiment, 0 hésite, -1 et -2 rejettent. Sur 25 lignes, ça se fait à la main ou dans un tableur. Un chiffre écrit dans le rapport sans calcul derrière ne vaut rien, on l'a vu avec Haiku.

Prompt 5, l'entretien

Choisis une personne qui rejette et recopie sa fiche, son post et son commentaire.

> Tu joues [PRÉNOM], [ÂGE] ans, [MÉTIER] à [VILLE]. [SITUATION] Caractère : [CARACTÈRE]. [HABITUDES] Tu as vu ce produit sur un réseau social : « [TON IDÉE, AVEC SON PRIX] ». Ton post : « [SON POST DU TOUR 1] ». Ton commentaire ensuite : « [SON COMMENTAIRE DU TOUR 2] » (avis [SON SCORE] sur une échelle de -2 à 2). Un enquêteur te demande : « Qu'est-ce qui te ferait changer d'avis et [ACHETER, T'ABONNER, PRENDRE RENDEZ-VOUS] ? » Réponds à l'oral, avec ta voix, en 35 mots maximum. Réponds uniquement par le texte de ta réponse.

Garder la même foule pour comparer deux IA

Pour que la comparaison veuille dire quelque chose, une seule chose doit changer : le modèle.

- Génère les personnes une fois et garde le fichier. Ne les régénère jamais entre deux modèles, sinon tu compares deux foules.
- Fixe les amis par une règle ou par le code, jamais par l'IA.
- Recolle les mêmes prompts, le même texte d'idée, les mêmes lots de 25.
- Change de modèle dans le sélecteur de claude.ai, ou avec l'option --model de Claude Code (sonnet, haiku, opus).
- Avant de conclure, relance deux fois le même modèle. Si deux lancements identiques s'écartent autant que deux modèles, l'écart ne veut rien dire. On ne l'a pas fait pour ce test : c'est sa première limite.

Dans Claude Code, pour 100 personnes

Ouvre Claude Code dans un dossier vide et colle ce prompt, suivi des cinq prompts ci-dessus. Les options citées sont celles de Claude Code 2.1.280, la version du 8 novembre 2026 ; tape claude --help si l'une manque chez toi.

> Dans ce dossier, écris un script Python simulate.py qui simule une foule avec Claude Code en mode non interactif. Chaque appel passe par subprocess avec la commande claude -p --safe-mode --model <modèle> --tools "" --output-format json --no-session-persistence, le prompt envoyé sur l'entrée standard, et s'enregistre dans journal/ (prompt, réponse, durée, coût indiqué). Étapes : 1) 100 personnes en 4 lots de 25 avec le prompt 1, une seule fois, dans personas.json ; 2) un graphe d'amis fixé par le code (chacun connaît ses deux voisins de chaque côté dans la liste, plus un lien tiré au hasard pour 35 % des personnes, graine 92), ajouté à personas.json ; 3)

tour 1 par lots de 25 avec le prompt 2 ; 4) tour 2 par lots de 25 avec le prompt 3, en donnant à chaque personne son avis du tour 1 et les posts de ses amis ; 5) rapport avec le prompt 4 ; 6) entretien avec le prompt 5, avec une personne qui rejette. Le script compte lui-même les scores (1 et 2 = aime, 0 = hésite, -1 et -2 = rejette) et les avis changés entre les deux tours, sans se fier au rapport. Le modèle se choisit en argument : python3 simulate.py run sonnet, puis python3 simulate.py run haiku sur le même personas.json. Voici les cinq prompts : [COLLE LES PROMPTS 1 À 5]

Le graphe d'amis de notre test, pour retrouver exactement les mêmes liens :

```
À COLLER

import random

def friends_graph(n, seed=92):
    rnd = random.Random(seed)
    g = {i: set() for i in range(1, n + 1)}
    for i in range(1, n + 1):
        for d in (1, 2):
            j = (i - 1 + d) % n + 1
            g[i].add(j); g[j].add(i)
    for i in range(1, n + 1):
        if rnd.random() < 0.35:
            j = rnd.randint(1, n)
            if j != i:
                g[i].add(j); g[j].add(i)
    return {i: sorted(v) for i, v in g.items()}
```

Les appels passent par ton abonnement et entament ta limite d'usage. Notre test complet valait 1,69 $ au tarif de l'API.

---

6. Les limites, dites franchement

Personne n'a montré qu'il prédit juste

Le README ne publie aucun taux de réussite, et le cadre de test sur des événements passés est une proposition de code toujours ouverte (n° 813, 15 septembre 2026). La FAQ officielle (ticket 726, 21 juillet 2026) écrit elle-même que MiroFish « must not be equated with a scientifically calibrated classical ABM » et que le code « does not automatically guarantee calibrated polarization, opinion drift, prediction confidence, or causal validity » (question 1). Pour le rapport : « provides no calibrated confidence score » (question 25).

SocietyBench (Wang, Bian, Li et Qi, arXiv 2608.04009, 4 août 2026) est le seul banc d'essai indépendant qu'on ait trouvé. Il demande à des modèles de prévoir la suite de cinq vrais événements de société, noms remplacés et dates décalées pour qu'ils ne puissent pas s'en souvenir. Répondre 50 % à tout donne 50 sur 100 ; le meilleur modèle, GPT-5.5, fait 75,0. MiroFish y tourne sur le modèle Doubao-Seed-2.0-Pro : 63,1, contre 63,4 pour le même modèle utilisé seul. Les auteurs : « None improves on the base: LangGraph loses 1.3 points overall,

AutoGen 2.9, and MiroFish 0.3. » Ce banc mesure des prévisions d'événements. On n'a trouvé aucun banc d'essai pour le test d'une idée de produit.

System	Cal.	Time	Cal.	Time	Cal.	Time	Cal.	Time	Cal.	Time	Cal.	Time	Overall
Doubao (base)	65.5	58.3	57.9	60.3	67.0	64.2	60.5	65.0	70.3	65.0	**64.3**	**62.6**	**63.4**
LangGraph	62.7	55.8	55.9	60.4	65.8	64.7	56.6	65.5	69.0	64.9	**62.0**$_{-2.3}$	**62.3**$_{-0.3}$	**62.1**$_{-1.3}$
AutoGen	59.2	57.2	52.6	61.4	63.9	64.7	50.3	64.5	66.6	65.0	**58.5**$_{-5.8}$	**62.5**$_{-0.1}$	**60.5**$_{-2.9}$
MiroFish	66.2	57.5	57.1	59.2	65.5	65.5	59.2	65.7	70.3	64.4	**63.7**$_{-0.6}$	**62.4**$_{-0.2}$	**63.1**$_{-0.3}$

Table 8 | Agent orchestration. Three agent frameworks running on the *Doubao* base model (first row); subscripts give the change relative to the base.

Tableau 8 de SocietyBench (arXiv 2608.04009, 4 août 2026, licence CC BY 4.0, extrait recadré) : sur le modèle Doubao, MiroFish fait 63,1 au total, le modèle seul 63,4. Les cinq paires de colonnes sont les cinq événements testés, les trois dernières la moyenne.

WPP Research a fait tourner MiroFish en mai 2026 sur une question de Bourse (le S&P 500), face à un simple appel à Gemini. Les deux se sont trompés, et l'auteur conclut : « The swarm added complexity without adding information. »

Deux lancements, deux foules

Le ticket 751 (2 août 2026, toujours ouvert) l'a mesuré : « re-running the same input builds a different crowd each time » : nouvelle ontologie, nouvelles entités, nouveaux personnages. Sur deux lancements du même texte, avec la même question d'entretien, 9,5 % puis 60,0 % des agents refusent de répondre. En forçant la température à 0 à trois endroits du code, trois lancements sur quatre ont produit les mêmes personnages à l'octet près. La FAQ le dit à sa façon : « Outputs vary with material, model, random seed, agent settings, and external APIs. »

Notre test va dans le même sens côté modèle : 36 personnes sur 100 changent de camp quand on passe de Sonnet à Haiku.

Ce que disent les chercheurs

- Wang, Morgenstern et Dickerson (Nature Machine Intelligence, 2025, 3 200 participants humains, 16 identités, 4 modèles) : « LLMs are likely to both misportray and flatten the representations of demographic groups ». Les modèles dépeignent mal les groupes et les rendent plus uniformes qu'ils ne sont.
- Bisbee et ses coauteurs (Political Analysis, mai 2024) ont fait noter des groupes politiques par ChatGPT dans la peau de personnes fictives, sur un « feeling thermometer » de 0 à 100. Les moyennes collent à la grande enquête électorale américaine, mais « there is less variation in responses than in the real surveys », et « the same prompt yields significantly different results over a 3-month period ».
- Ye, Cao, Chen et Ferrara (arXiv 2605.18890, 17 mai 2026) : de petits changements dans la façon d'écrire les personnages ou les consignes font bouger un taux de coopération de 76 points, et « the same perturbation that produces the 76 pp shift in one frontier model only shifts another by 1 pp ».
- Le papier d'OASIS, le moteur de MiroFish (Yang et coauteurs, arXiv 2411.11581, novembre 2024), trouve des agents « more susceptible to the herd effect compared to humans ». Les revirements au contact des autres sont donc à prendre avec des pincettes.

Les limites de notre propre test

Un seul lancement par modèle, une seule idée, des prompts écrits par nous. On n'a pas relancé deux fois le même modèle, donc on ne sait pas quelle part de l'écart entre Sonnet et Haiku vient du modèle et quelle part vient du hasard. Et nos 100 personnes ont été écrites par la même IA que celle qui les joue.

---

7. T'en servir comme thermomètre

Pour une PME ou un indépendant, une foule simulée sert à préparer le terrain avant de parler à de vrais clients. La conversation avec eux reste à faire.

Ce que tu peux en tirer

- Une liste d'objections, avec des mots de tous les jours. Dans notre test : le prix livraison comprise (« 48 € »), les gamelles maison, le repas pris dans le camion ou sur le chantier, les horaires décalés. Chacune devient une question à poser à de vraies personnes.
- Des groupes à aller voir en premier. Ici, les soignantes libérales et les cadres débordés. C'est une piste, rien de plus.
- Une comparaison entre deux versions sur la même foule. Par exemple 44,90 € plus 3,90 € de livraison contre 48,80 € livraison comprise, ou une livraison au bureau seulement. On ne l'a pas testé : lance chaque version deux fois et ne retiens l'écart que s'il dépasse celui entre deux lancements de la même version.
- Des questions d'entretien. La réponse de Nadia (15 à 20 € la semaine, livraison gratuite, des portions pour les enfants) te dit quoi demander à de vraies mères qui travaillent de nuit. Elle ne te donne pas leur prix.

Ce que tu ne peux pas en tirer

- Un taux de conversion. « 25 aiment sur 100 » ne veut pas dire que 25 % achèteront : Haiku et Sonnet ne sont même pas d'accord entre eux, et la recherche montre des foules simulées plus uniformes que les vraies.
- Un prix. Les 15 à 20 € de Nadia sortent d'un rôle joué par une IA.
- La taille de ton marché, ou la décision de lancer. Personne n'a payé, personne n'a rien risqué.

Confirmer avec de vraies personnes

Quatre façons peu coûteuses, à adapter à ton activité :

1. Parle à 5 à 10 personnes du groupe qui a le plus aimé dans la simulation, et demande-leur ce qu'elles ont fait, pas ce qu'elles feraient : ce qu'elles ont mangé à midi la semaine dernière, combien elles ont dépensé, ce qui les a agacées. C'est la méthode du livre The Mom Test de Rob Fitzpatrick (2013).

2. Pose les quatre questions de prix de Van Westendorp (1976) à 20 ou 30 personnes de ta cible : à partir de quel prix c'est trop cher pour que tu l'achètes, cher mais tu y réfléchis, une bonne affaire, trop bon marché pour que ce soit sérieux. Tu obtiens une fourchette tirée de vraies réponses.

3. Mets en ligne une page avec le vrai prix et un bouton « je réserve ma première semaine », partage-la dans deux ou trois groupes locaux, et compte les clics par rapport aux visites. Un clic engage plus qu'un « j'adore » en commentaire.

4. Demande de l'argent. Dix personnes qui paient une semaine d'avance, servies à la main avec un traiteur du coin avant d'écrire la moindre ligne de code, valent plus que mille agents. Un oui qui ne coûte rien ne compte pas.

Garde les objections de la simulation à côté des réponses réelles. Celles que la foule simulée n'avait pas vues te disent jusqu'où lui faire confiance la prochaine fois.

---

8. Où lire et vérifier toi-même

- Le dépôt : github.com/666ghj/MiroFish, avec le README en chinois et le fichier .env.example.
- La FAQ officielle (ticket 726), à lire avant d'installer : coûts (question 22), Zep (question 11), sécurité (question 16), faux sites (question 2).
- Le ticket sur la reproductibilité et celui sur le nombre d'agents.
- La démo en ligne figée sur l'université de Wuhan, sans IA branchée.
- SocietyBench (arXiv, 4 août 2026), tableau 8 page 11.
- Le test de WPP Research (mai 2026).
- OASIS, le moteur de simulation de CAMEL-AI.
- La CNIL : questions-réponses sur l'utilisation d'un système d'IA générative (18 juillet 2024).

Sources

Pages relues et chiffres relevés le 8 novembre 2026.

- MiroFish : dépôt et API GitHub (étoiles, forks, licence, commits), FAQ n° 726 (21 juillet 2026, vérifiée par l'équipe le 23 juillet), tickets 751 et 838, code de la branche main (fichiers ontology_generator.py, oasis_profile_generator.py, config.py, report_agent.py).
- Presse : Science and Technology Daily (9 mars 2026), Xinmin Evening News (11 mars 2026), ChinaVenture (13 mars 2026), Wenhui Daily (18 mars 2026).
- Classement GitHub Trending : archives Wayback du 7 mars 2026 et du 16 mars 2026.
- Shanda : Hurun China Rich List 2026 (27 octobre 2026).
- Études : SocietyBench (licence CC BY 4.0), WPP Research, Wang, Morgenstern, Dickerson (Nature Machine Intelligence, 2025), Bisbee et al. (Political Analysis, 2024), Ye, Cao, Chen, Ferrara (2026), OASIS (2024).
- Prix : Alibaba Cloud Model Studio (7 novembre 2026) et ses régions (6 novembre 2026), Zep, taux de référence de la BCE du 6 novembre 2026.
- Sécurité et droit : CVE-2026-7042 (NVD, 26 avril 2026), CNIL (18 juillet 2024), RGPD sur EUR-Lex, licence AGPL-3.0.
- Notre test : dossier reels-out/92-mirofish-simulateur/test (simulate.py, personas.json, run-sonnet, run-haiku, journal de chaque appel).