Eigenflow

GUIDE · CLAUDE CODE

Arena pour Claude Code

Installer la skill qui fait s'affronter tes sous-agents, la lancer, ce qu'elle coûte, et notre vrai tournoi du 21 octobre 2026 : 16 agents, 4 tours, 79 contre 60.

PAR

Abyane Nabeebaccus

INSTAGRAM

@abyane_nb

ÉDITION

novembre 2026

Tu as commenté ARENA, voici le guide. Arena est une skill gratuite pour Claude Code, écrite par Jake Schincariol et publiée sous licence MIT le 27 septembre 2026. Quand une réponse de Claude ne te plaît pas, elle lance des dizaines de sous-agents sur la même demande, les fait s'attaquer deux par deux, et te rend la seule réponse encore debout à la fin.

On l'a fait tourner le 21 octobre 2026 sur une demande toute simple : « Écris un mail de prospection pour décrocher mon premier client. » Ce guide raconte ce test avec ses chiffres et des extraits tirés des fichiers, puis explique comment installer Arena, la lancer, ce qu'elle coûte, quand elle vaut le coup et quand elle ne sert à rien. On a tout gardé : la réponse normale, les seize réponses de départ, chaque attaque, chaque défense et chaque verdict.

À savoir avant de lire

- Arena ne tourne que dans Claude Code, parce qu'elle lance ses concurrents avec l'outil Agent de Claude Code. Il faut aussi Python 3.8 ou plus récent. Rien à installer avec pip.
- La skill est gratuite, les tokens sont à toi. Par défaut : 100 sous-agents et 595 appels. Avec --quick : 16 sous-agents et 91 appels. Notre --quick a coûté 24,30 $ au prix de l'API (environ 21,70 €), payés ici par l'abonnement, et a duré 32 min 34 s.
- « Cent versions de Claude », ce sont cent sous-agents du même modèle. Une carte de stratégie les distingue. Les juges sont Claude, eux aussi.
- Notre résultat : 16 agents, 4 tours, le champion A015 a gagné ses 4 combats. Le juge final, à l'aveugle, a mis 79 sur 100 à sa réponse et 60 à la réponse normale. La gagnante fait 3 950 mots, la normale 385.
- Les phrases marquées « testée » ont été dites telles quelles le 21 octobre 2026 (Claude Code 2.1.280, modèle Opus, mode restreint). Les autres portent la mention « non testée ».
- Le dépôt n'a qu'un seul commit (df07b8e, 27 septembre 2026). Le 10 novembre, il comptait 1 406 étoiles. On a lu son script avant le test : bibliothèque standard de Python, écriture dans le seul dossier .arena/, aucun accès réseau.

1. Ce que fait Arena

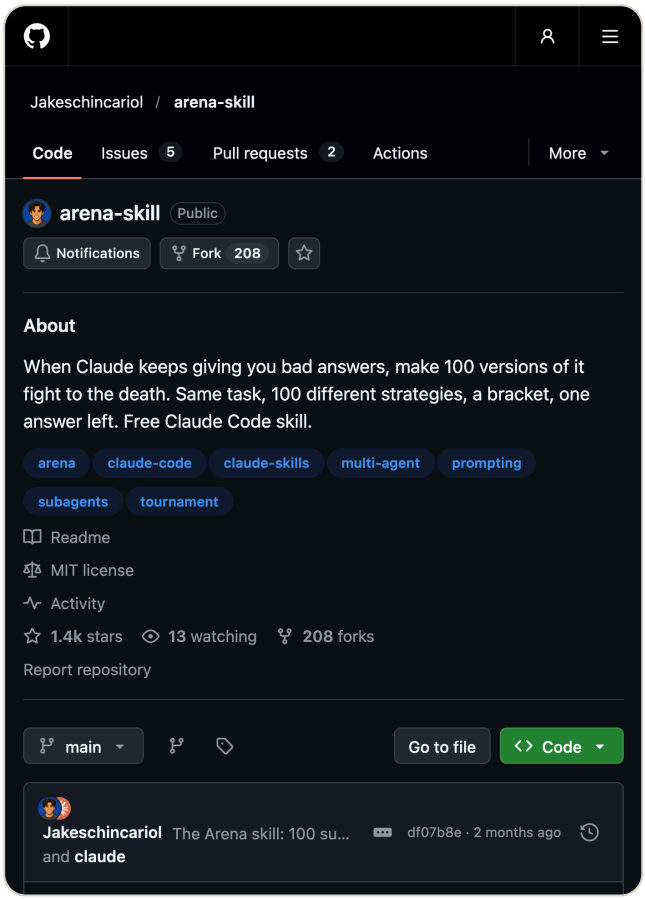
Le README du dépôt la présente en une phrase : « When Claude keeps giving you bad answers, make 100 versions of it fight to the death. » Si Claude continue de te donner de mauvaises réponses, fais se battre cent versions de lui, jusqu'à la mort.

Concrètement, tu tapes /arena après une réponse qui te déçoit. La skill écrit ta demande dans un fichier, la donne au mot près à 100 sous-agents, et donne à chacun une carte différente : une façon de raisonner, une méthode de travail et une stratégie. Chacun écrit sa réponse. Puis ils sont

appariés. Chacun attaque la réponse de l'autre, défend la sienne et la corrige, et un juge note le match sur un barème écrit. Le perdant sort. 100 deviennent 50, puis 25, 13, 7, 4, 2 et 1.

Ce qui te revient : la réponse qui a survécu, les attaques qu'elle a encaissées en route, sa carte et le nombre de tours. Si tu étais parti d'une réponse refusée, un dernier juge compare les deux sans savoir laquelle est laquelle, et la skill te donne les notes, même quand l'ancienne gagne.

Arena ne touche pas à tes fichiers. Chaque sous-agent écrit dans un dossier .arena/ créé là où tu lances Claude Code. Si la tâche porte sur du code, les changements reviennent sous forme de diff ou de fichiers complets dans la réponse gagnante, et la skill te demande avant de les appliquer.



La page du dépôt Jakeschincariol/arena-skill sur GitHub, capturée le 10 novembre 2026 : licence MIT, 1,4k étoiles, et un seul commit, df07b8e.

---

2. Quand t'en servir, et quand passer ton chemin

Les bons cas

- Une réponse qui compte et qui peut se faire de plusieurs façons : un mail important, un plan de lancement, le texte d'une page de prix, un choix entre deux architectures, un prompt que tu vas réutiliser cent fois. Arena sert quand plusieurs approches se défendent et qu'on gagne à les comparer.
- Une réponse que tu viens de refuser. Arena la prend comme réponse à battre et te dit franchement si elle reste meilleure que la gagnante.
- Le README donne lui-même la règle : --quick pour les choses de tous les jours, les 100 « pour la réponse qui compte ».

Les mauvais cas

- Une question de fait : une date, un prix, un article de loi. Les sous-agents ont tous le même savoir, et les juges vérifient la logique des réponses, pas le monde. Dans notre test, ils n'avaient même pas accès au web.
- Une demande floue. Le README le dit sans détour : « Vague task in, 100 flavours of vague out. » Notre test en est un bon exemple. Faute de connaître le métier de l'utilisateur, chaque sous-agent a dû couvrir tous les métiers, et la réponse a enflé.
- Quand tu veux une réponse courte. Chez nous, la réponse du champion faisait 1 039 mots au départ, puis 1 733, 2 302, 3 020 et enfin 3 950 mots après la finale. Le barème ne récompense pas la longueur, mais chaque défense ajoute des corrections.
- Quand tu es pressé. La réponse normale a pris 14 secondes. Le tournoi en --quick, 32 minutes et 34 secondes.
- Quand ton quota est déjà serré. Le mode par défaut, c'est 595 appels de sous-agents.

---

3. Installer Arena : les trois façons du README

Le README propose trois chemins. Aucun n'a été testé tel quel chez nous (voir plus bas ce qu'on a fait), mais les trois mènent au même dossier skills/arena.

Coller le lien dans Claude Code (non testée)

C'est la méthode du reel. Ouvre Claude Code et colle ces deux lignes, telles que le README les donne :

```
À COLLER

https://github.com/Jakeschincariol/arena-skill

Install this skill, then confirm /arena works.
```

Claude choisit alors lui-même comment installer. Regarde ce qu'il fait : s'il passe par le système de plugins, la commande s'appellera /arena-skill:arena et pas /arena.

Copier le dossier toi-même (non testée)

Dans un terminal :

```
À COLLER

git clone https://github.com/Jakeschincariol/arena-skill.git
cp -r arena-skill/skills/arena ~/.claude/skills/
```

La commande sera /arena, dans tous tes projets. Pour un seul projet, copie le même dossier dans le .claude/skills/ de ce projet.

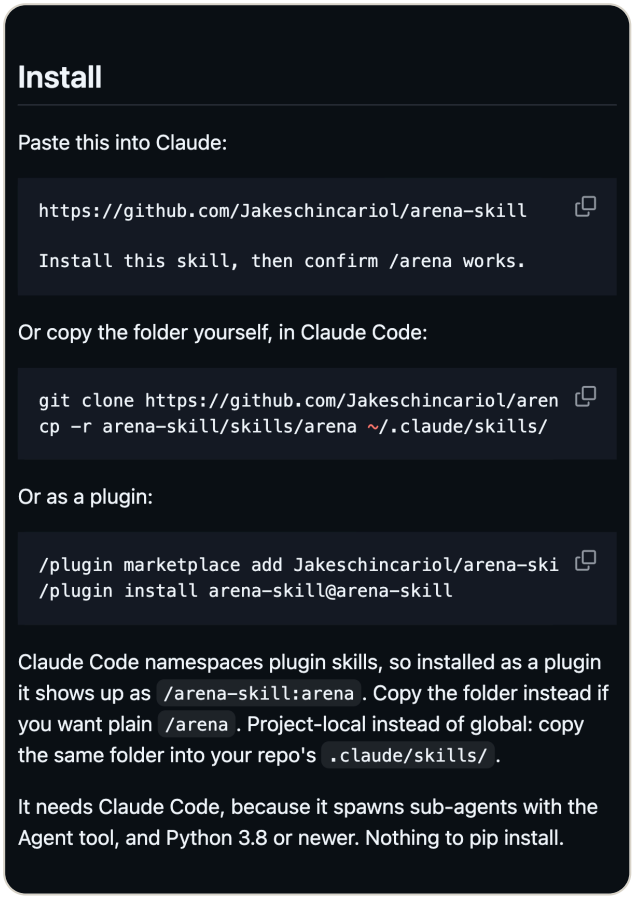
L'installer comme plugin (non testée)

Dans Claude Code :

```
À COLLER

/plugin marketplace add Jakeschincariol/arena-skill
/plugin install arena-skill@arena-skill
```

Claude Code range les skills des plugins sous le nom du plugin. Installée ainsi, la commande devient /arena-skill:arena.



La partie « Install » du README, capturée le 10 novembre 2026 : la phrase à coller dans Claude, la copie du dossier, et l'installation en plugin.

Ce qu'on a fait, nous

On a copié SKILL.md, bracket.py, rubric.md, strategies.json et le fichier du plugin dans un dossier, chargé seul avec l'option --plugin-dir de Claude Code. Dans notre test, la commande s'appelait donc /arena-skill:arena.

Avant de lancer quoi que ce soit, on a lu bracket.py, le script qui tient les comptes du tournoi. Il n'importe que des modules de la bibliothèque standard de Python (argparse, json, math, os, random, re, sys, time), n'écrit que dans .arena/, n'appelle pas le réseau et ne supprime rien : au pire, il renomme un verdict illisible en .unreadable. La skill ne choisit aucun modèle, elle prend celui de ta session. Relu le 10 novembre : le dépôt n'a pas bougé depuis son unique commit.

Les prérequis

- Claude Code, parce que la skill lance ses concurrents avec l'outil Agent. Elle ne marche pas dans l'appli de chat.
- Python 3.8 ou plus récent.
- Une version récente de Claude Code. Jusqu'à la version 2.1.224, une session plafonnait par défaut à 200 sous-agents (variable CLAUDE_CODE_MAX_SUBAGENTS_PER_SESSION, supprimée depuis, d'après la documentation relue le 10 novembre). Le mode par défaut d'Arena en demande 595.

---

4. Lancer Arena

Les quatre exemples du README, traduits :

```
À COLLER

/arena
/arena --quick écris le titre de notre page de prix
/arena --agents 32 corrige le test instable dans tests/test_api.py
/arena --seed 7 planifie ma semaine de lancement, j'ai 6 heures par jour
```

Les réglages :

- --agents N : le nombre de concurrents. 100 par défaut.
- --quick : 16 concurrents. Le README l'appelle « le réglage de tous les jours ».
- --seed S : la même graine redonne les mêmes cartes et le même tableau. Pas les mêmes réponses, le modèle n'étant pas déterministe. Sans graine, la skill en tire une et la note (la nôtre : 422633).
- --wave W : le nombre de sous-agents lancés en même temps. 10 par défaut. Le README conseille de ne pas y toucher sans avoir relevé la limite de Claude Code.

Avant de démarrer, la skill annonce la taille du tournoi en une ligne, par exemple « 100 agents, 7 rounds, 595 sub-agent calls ».

Sans tâche écrite, Arena reprend ta dernière demande

Tapé seul, /arena prend comme tâche ta dernière demande de la conversation, et la dernière réponse de Claude comme réponse à battre. C'est ce qu'on a fait. Tout ce que tu écris après les réglages devient la tâche.

Les sous-agents, eux, ne voient pas ta conversation. Ils ne connaissent que le fichier de tâche écrit par Claude, rangé dans .arena/ (le chemin exact est .arena/<run>/task.md). D'après SKILL.md, Claude doit y recopier toutes les contraintes que tu as données dans la conversation, et ce qui t'a déplu dans la réponse à battre, avec tes mots. Si tu n'as rien dit, il n'y a rien à recopier. Notre fichier de tâche disait, traduit : « L'utilisateur n'a donné aucune autre information : ni son métier ou son service, ni ses clients cibles, ni son secteur, ni son prix », puis « Il n'a pas dit ce qui lui déplaisait. »

Donc, avant de taper /arena, écris dans un message normal ton métier, ta cible, ton prix, et ce qui cloche dans la réponse. Puis tape /arena --quick seul.

Quand Claude la déclenche tout seul

La skill peut aussi partir sans la commande, quand tu dis « mauvaise réponse » ou « fais-les s'affronter ». Dans ce cas, SKILL.md lui demande de te poser la question une fois avant de dépenser quoi que ce soit, avec trois choix : l'arène complète (100 agents, 595 appels), --quick (16 agents, 91 appels) ou un simple nouvel essai.

Les permissions

Chaque sous-agent écrit un fichier dans .arena/. En mode manuel, c'est une demande d'accord par fichier, des centaines sur un grand tournoi. Le README conseille le mode accept-edits, avec Shift+Tab.

Deux précisions tirées de la documentation de Claude Code, relue le 10 novembre 2026. Depuis la version 2.1.283, une session interactive démarre en mode auto, et Shift+Tab fait tourner les modes. Quand la barre d'état affiche « accept edits on », les écritures de fichiers dans ton dossier passent sans demande. Mais les commandes du terminal non prévues par ce mode demandent toujours ton accord, et Arena appelle bracket.py avec python3 à chaque étape.

Notre test réglait les deux d'avance : mode acceptEdits, et commandes python3 autorisées. Ces deux options ont servi telles quelles le 21 octobre, avec claude -p :

```
À COLLER

claude --permission-mode acceptEdits --allowedTools "Bash(python3:*)"
```

Même comme ça, nos réglages ont refusé 31 actions pendant le tournoi, surtout des comptages de mots avec wc, sed ou awk. Le tournoi est allé au bout quand même. Dans une session en accept-edits, ces commandes te seraient demandées. En mode auto, c'est un classificateur qui tranche, et on ne l'a pas essayé avec Arena.

Lance toujours Arena depuis le dossier où tu veux voir apparaître .arena/. Toutes les commandes du tournoi partent de là.

---

5. Comment se déroule le tournoi

Les cartes

Le fichier strategies.json contient 15 façons de raisonner, 12 méthodes et 12 stratégies, soit 15 × 12 × 12 = 2 160 cartes. Quelques exemples traduits : premiers principes, inversion, analogie, adversaire, à rebours, probabiliste ou panel d'experts pour le raisonnement ; plan d'abord, trois brouillons, brouillon puis critique puis réécriture, la plus petite version d'abord pour la méthode ; le plus simple qui marche, rigueur maximale, l'utilisateur d'abord, les cas limites d'abord pour la stratégie.

Le distributeur ne donne jamais deux fois la même carte, répartit chaque élément aussi également que possible, et jusqu'à 144 agents, deux agents ne partagent jamais deux éléments sur trois. Le fichier se modifie librement.

La carte de notre champion, A015, avec les consignes du fichier traduites :

- Raisonnement « Adversaire » : imagine le relecteur expert le plus dur qui essaie de démolir la réponse, écris ses attaques d'abord et construis la réponse à partir d'elles.
- Méthode « La plus petite version d'abord » : produis la plus petite version complète qui marche, puis étends-la une étape à la fois, sans jamais la casser en route.
- Stratégie « L'utilisateur d'abord » : pars de la personne qui va s'en servir, de sa situation, de son niveau et de ce qu'elle fera ensuite.



La carte d'A015 dans le reel : Adversaire, La plus petite version d'abord, L'utilisateur d'abord. Les noms sont ceux de strategies.json, traduits.

Un match en quatre temps

1. L'attaque. Chacun attaque la réponse de l'autre à travers sa propre carte : ce qui est faux, l'exigence oubliée, le cas précis qui la casse. Sept attaques au plus, classées FATAL, MAJOR ou MINOR, chacune avec la citation visée.

2. La défense. Chacun répond à chaque attaque reçue : CONCEDE s'il a tort, REBUT s'il peut prouver le contraire. Le brief est clair, une réfutation qui se contente d'insister compte comme une concession. Puis il réécrit sa réponse en entier, corrections comprises.

3. Le jugement. Un sous-agent juge, à part, lit les deux réponses corrigées, vérifie lui-même chaque attaque et note les deux sur le barème. Il ne voit jamais les cartes.

4. Le calcul. bracket.py fait l'addition pondérée et le meilleur total passe. Une réponse avec une faille fatale vérifiée ne peut pas battre une réponse sans. En cas d'égalité, c'est celle qui a le moins d'attaques encore debout qui l'emporte, puis la meilleure note de justesse.

Le barème

Le juge note chaque critère de 0 à 10, puis le script pondère :

- Justesse : 30 points sur 100. Rien de faux, pas d'erreur de logique.
- Complétude : 25 points. Chaque exigence écrite dans la tâche est remplie.
- Solidité face aux attaques du match : 20 points.
- Précision : 15 points. On peut s'en servir tout de suite, sans deviner.
- Clarté : 10 points. Facile à lire, à une longueur qui colle à la tâche.

Le fichier rubric.md précise ce que le juge ne récompense pas : la longueur, l'assurance du ton, la méthode employée, et les phrases qui vantent la qualité de la réponse. Il se modifie lui aussi.

Exemptions et vagues

Quand le nombre de survivants est impair, un agent passe au tour suivant sans combattre, et jamais deux fois le même tant qu'un autre attend son tour. En mode 100, ça arrive aux tours 3, 4 et 5 (25, 13 et 7 survivants). Le champion d'un tournoi à 100 joue donc 6 ou 7 matchs, pas 99. Les autres ont été sortis par d'autres concurrents.

La skill lance les sous-agents par vagues de 10. Un tournoi à 100, c'est 70 vagues. La session principale ne lit jamais les réponses : tout l'état du tournoi tient dans un fichier arena.json, et si la conversation est compactée en route, bracket.py next reprend là où ça s'était arrêté.

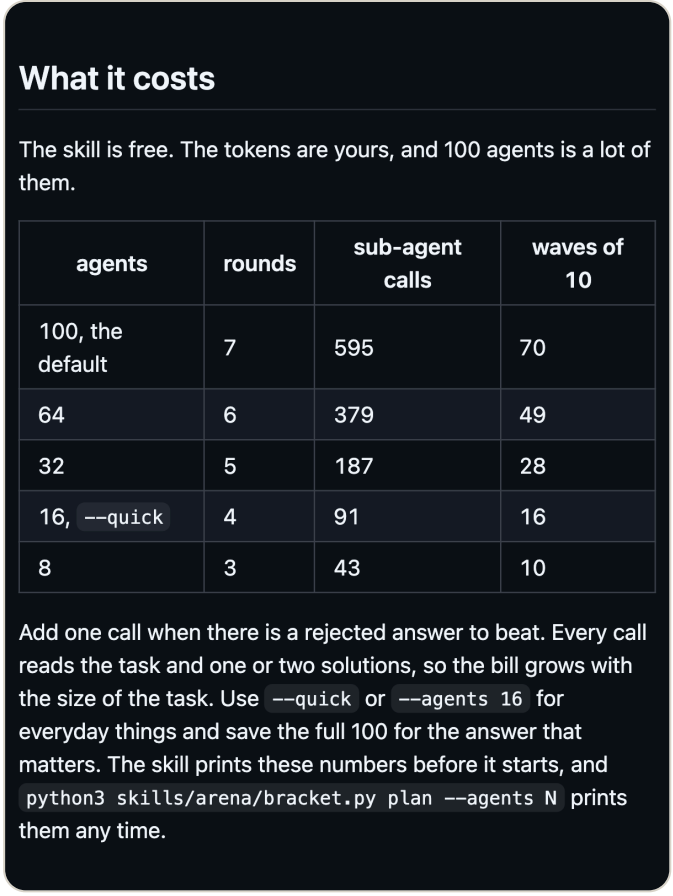
---

6. Ce que ça coûte

Le README le résume ainsi : « The skill is free. The tokens are yours, and 100 agents is a lot of them. » La skill est gratuite, les tokens sont à toi. Son tableau :

- 100 agents, le défaut : 7 tours, 595 appels de sous-agents, 70 vagues de 10.
- 64 agents : 6 tours, 379 appels, 49 vagues.
- 32 agents : 5 tours, 187 appels, 28 vagues.
- 16 agents, --quick : 4 tours, 91 appels, 16 vagues.
- 8 agents : 3 tours, 43 appels, 10 vagues.

Ajoute un appel quand il y a une réponse refusée à battre. Chaque appel relit la tâche et une ou deux réponses, donc la facture grossit avec la taille de la tâche, et avec la longueur des réponses qui enflent de tour en tour.



What it costs

The skill is free. The tokens are yours, and 100 agents is a lot of them.

agents	rounds	sub-agent calls	waves of 10
100, the default	7	595	70
64	6	379	49
32	5	187	28
16, `--quick`	4	91	16
8	3	43	10

Add one call when there is a rejected answer to beat. Every call reads the task and one or two solutions, so the bill grows with the size of the task. Use `--quick` or `--agents 16` for everyday things and save the full 100 for the answer that matters. The skill prints these numbers before it starts, and `python3 skills/arena/bracket.py plan --agents N` prints them any time.

Le tableau des coûts du README, capturé le 10 novembre 2026.

Ce qu'on a mesuré en --quick, le 21 octobre :

- 92 sous-agents lancés : les 91 du tournoi et le juge final. Aucun échec.
- 32 min 34 s du début à la fin.
- 24,30 $ affichés par Claude Code en équivalent API, soit environ 21,70 € au cours de la BCE du 9 novembre 2026. La réponse normale avait coûté 0,04 $. Sur notre abonnement, rien n'a été payé en plus. On n'a pas mesuré la part de la limite d'usage que ça a mangé.
- 630 193 tokens de sortie : 587 009 par claude-opus-5-5 (93 %), 42 626 par claude-opus-4-8, et le reste par claude-opus-5 et claude-haiku-4-5. La skill ne choisit pas de modèle, c'est Claude Code qui a réparti.

On n'a pas lancé le mode 100. On ne donne donc pas de chiffre pour lui.

---

7. Notre test du 21 octobre, pas à pas

Le cadre

Un dossier vide sur un Mac. Claude Code 2.1.280, modèle Opus, en mode restreint : les réglages, hooks et skills de nos projets étaient ignorés, aucun serveur MCP n'était branché, et Arena était la seule skill chargée. Outils permis : lecture, écriture, modification et recherche de fichiers, le terminal limité à python3, et l'outil Agent.

On a d'abord posé la demande, puis tapé, dans la même session :

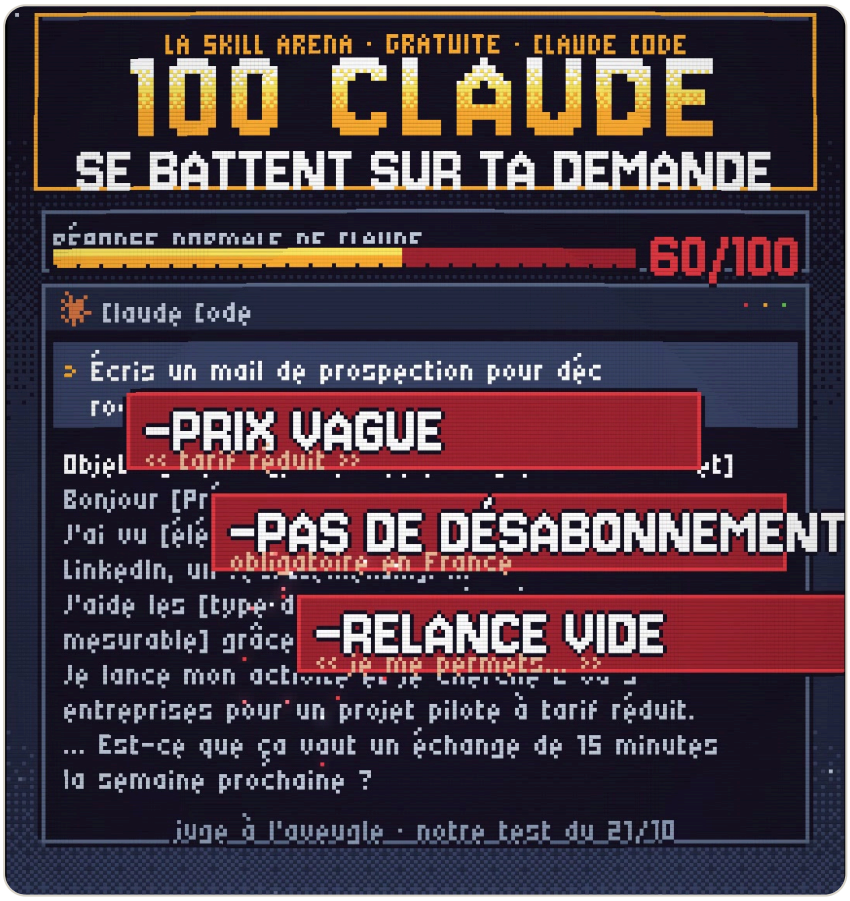
```
À COLLER

/arena-skill:arena --quick
```

La réponse normale : 385 mots, 14 secondes

Claude a rendu un modèle à trous, propre, avec un objet, une accroche personnalisée et une demande de 15 minutes. Deux phrases lui ont coûté cher devant le juge final. Dans le mail : « Je lance mon activité et je cherche 2 ou 3 entreprises pour un projet pilote à tarif réduit. » Et dans la relance : « Bonjour [Prénom], je me permets de remonter mon message. »

Le juge final, qui ne savait pas laquelle des deux était la réponse normale, l'a résumée ainsi (traduit) : « un modèle générique avec un « tarif réduit » vague, pas de ligne de désabonnement et une relance vide ». À sa décharge, la relance disait quand même « dites-le-moi simplement et je ne vous relancerai plus ». Le premier mail, lui, n'offrait aucun moyen de refuser.



Le début du reel : la réponse normale de Claude, notée 60 sur 100 par le juge final, et ses trois défauts relevés dans le verdict.

Le tournoi

Seize agents, quatre tours, aucune exemption (16 se divise par deux jusqu'au bout). Le parcours du champion :

- Tour 1 : A015 bat A016, 86 à 82,5.
- Tour 2 : A015 bat A010, 88,5 à 84.
- Demi-finale : A015 bat A012, 91 à 87.
- Finale : A015 bat A014, 85 à 83. L'autre demi-finale avait vu A014 battre A009, 89 à 88.

Sur les 15 matchs, les concurrents se sont lancé 176 attaques : 98 majeures, 78 mineures, aucune fatale. Ils en ont concédé 174 et réfuté 2. A015 a reçu 24 attaques, 6 par match, et les a

toutes concédées et corrigées. On s'attendait à plus de résistance. Dans les faits, chaque match ressemble à une relecture serrée où presque tout est concédé.

Un détail : A014 et A015 tiraient tous les deux le raisonnement « Adversaire ». Le tableau évite d'opposer deux agents qui raisonnent pareil, mais en finale il ne restait plus qu'eux.



La fin du tournoi dans le reel : A015 a gagné ses 4 combats, la finale 85 à 83, en 33 minutes (32 min 34 s mesurées).

Une vraie attaque, et ce qu'A015 en a fait

La réponse de départ d'A015 commençait par : « Je ne connais pas ton métier, donc je te donne un mail qui marche pour n'importe quelle activité ». Au tour 1, A016 l'a attaquée là-dessus, en attaque majeure. Son titre, traduit : « Affirmation fausse, le modèle « marche pour n'importe quelle activité » : il ne marche que si le problème du prospect se voit de l'extérieur. » Son contre-exemple : une comptable indépendante qui regarde le site d'un plombier ne peut pas « remarquer » que ses factures partent en retard. A016 estimait que 40 à 60 % des métiers possibles tombaient dans ce cas.

A015 a concédé. Il a retiré « n'importe quelle activité », ajouté une deuxième façon d'ouvrir le mail pour les métiers où rien ne se voit, et un deuxième exemple rempli. Dans la réponse finale, cela donne trois ouvertures, et cet exemple pour une assistante administrative : « J'ai vu sur votre page Facebook que vous venez d'embaucher un deuxième plombier. Quand l'activité grossit, beaucoup d'artisans se retrouvent avec des factures en retard et des clients qui oublient de payer. »

r1/r1-m07.a016.attack.md · écrit par A016, contre A015 (extrait)

```
ATTACK 1 [MAJOR] False claim that the template "marche pour n'importe quelle activité":
it only works when the prospect's problem can be seen from outside
Where: Intro "je te donne un mail qui marche pour n'importe quelle activité" (...)
Problem: The task says nothing is known about the user's job. The whole email depends on
 spotting a specific flaw in the prospect's public presence (...) Counterexample: a
freelance bookkeeper looking at a plumber's website cannot "remarquer" that their
invoicing is late, so the first line, the free idea and the subject cannot be filled in
honestly. (...)
```

r1/r1-m07.a015.defense.md · la réponse d'A015 (extrait)

```
ATTACK 1: CONCEDE.
The "j'ai remarqué que..." opening only works when the problem can be seen from outside.
 I removed the "n'importe quelle activité" claim and added a second opening for jobs
where nothing is visible (a trigger such as a new opening, hiring or growth, or a direct
 question). I also added a second filled example for a bookkeeper, so the hard case is
shown and not just the easy web-design one.
```

L'attaque d'A016 et la concession d'A015, dans les fichiers du match r1-m07 (extraits, en anglais comme tous les échanges internes du tournoi).

Dans l'autre sens, A015 a frappé A016 sur sa garantie « Si le résultat ne vous convient pas, vous ne payez pas » : dans son exemple, le restaurant pouvait garder les 12 photos déjà en ligne sans rien payer. A015 avait la même garantie dans son propre mail, et A016 la lui a reprochée aussi, en attaque mineure. Les deux ont corrigé. La version finale d'A015 dit : « Avant que vous ne payiez quoi que ce soit, je vous montre gratuitement une maquette : si elle ne vous convient pas, on en reste là. »

Le verdict final, à l'aveugle

Le juge final a reçu deux réponses nommées X et Y, sans savoir laquelle sortait du tournoi. Ses notes sur 10 : X (la gagnante) 8 en justesse, 9 en complétude, 9 en précision, 7 en solidité, 5 en clarté, soit 79 sur 100. Y (la réponse normale) 6, 7, 5, 4 et 9, soit 60. Sa raison, traduite : X donne « un mail de prospection en français qu'on peut vraiment envoyer, avec un vrai prix, une ligne de désabonnement obligatoire, des exemples remplis et un chemin de la réponse au devis signé ». Et il ajoute : « Le principal défaut de X est sa longueur excessive. »

final.verdict.json · juge final à l'aveugle (X gagnante, Y normale)


```
{
  "scores": {
    "X": {"correctness": 8, "completeness": 9, "specificity": 9, "robustness": 7,
"clarity": 5, "fatal": false},
    "Y": {"correctness": 6, "completeness": 7, "specificity": 5, "robustness": 4,
"clarity": 9, "fatal": false}
  },
  "winner": "X",
  "reason": "X gives a French prospecting email that can actually be sent, with a real
price, a legally required opt-out, worked examples and a path from reply to signed
quote, while Y is a generic template with a vague 'tarif réduit', no opt-out line and an
 empty follow-up; X's main weakness is its excessive length.",
  "fixed": [
    "concrete price method instead of vague 'tarif réduit'",
    "opt-out / address-source PS required for French B2B email prospecting",
    "warns that cold-emailing private individuals needs prior consent and offers
alternatives",
    "three fully filled-in example emails",
    "follow-up that adds value instead of 'je me permets de remonter'",
    "low-effort one-word CTA plus steps from 'oui' to deposit and signed quote",
    "opening variants for cases where no visible problem exists"
  ]
}
```


Le verdict du juge final, tel qu'il est dans le fichier.

Ce que contient la réponse gagnante

Sept parties numérotées : le mail avec ses trois ouvertures, une méthode pour fixer son prix en 15 minutes (avec un plancher de 30 € de l'heure), trois exemples remplis, une relance, le chemin du « oui » au premier paiement, un message pour son entourage, et un passage sur les clients particuliers. Quelques phrases, telles quelles :

- « Je vous propose donc de refaire votre page d'accueil pour 250 €. »
- « PS : j'ai trouvé votre adresse sur votre site. Si vous ne souhaitez pas recevoir d'autre message de ma part, répondez simplement « non merci » et je ne vous écrirai plus. »
- « Étape 3. Envoie un devis le jour même. » Parmi ses mentions : « 30 % d'acompte à la signature, le reste à la livraison ».
- « Une seule relance suffit. Elle doit apporter quelque chose de nouveau. »

On a vérifié la partie juridique sur la page de la CNIL consacrée à la prospection par courrier électronique (mise à jour le 10 juin 2026, relue le 10 novembre). Entre professionnels, l'envoi sans accord préalable est possible quand l'offre a un rapport avec la profession de la personne, à condition de l'informer et de la laisser s'opposer. Et la CNIL écrit : « chaque sollicitation doit obligatoirement permettre à la personne concernée de prendre connaissance de l'identité de l'organisation qui l'émet ainsi que d'exprimer, si elle le souhaite et par un moyen simple, son refus de recevoir de nouvelles sollicitations ». Vers des particuliers, il faut leur accord préalable. La réponse gagnante dit la même chose, et ce guide ne vaut pas avis juridique.

8. Les phrases à dire à Claude

La demande de notre test (testée)

> Écris un mail de prospection pour décrocher mon premier client.

Lancer Arena sur la réponse que tu viens de refuser (testée)

> /arena --quick

Testée sous le nom /arena-skill:arena --quick, celui que prend la skill chargée en plugin. Copiée dans ~/.claude/skills/, elle s'appelle /arena.

Dire d'abord ce qui ne va pas (non testée)

À envoyer comme un message normal, juste avant /arena --quick, pour que le fichier de tâche le contienne :

> Ta réponse ne me va pas : le prix est flou, il manque la ligne pour se désabonner et la relance n'apporte rien. Je suis [ton métier], je vise [ton type de clients], mon prix est de [ton prix].

Donner la tâche directement (non testée)

> /arena --quick Écris un mail de prospection de 150 mots maximum pour [ton métier] qui vise [ton type de clients], avec un prix fixe et une ligne pour se désabonner.

Choisir la taille du tournoi (non testée)

> /arena --agents 32 [ta demande]

Rejouer le même tirage (non testée)

> /arena --seed 7 [ta demande]

Raccourcir la réponse gagnante (non testée)

Le défaut relevé par le juge final, c'était la longueur. Une fois le tournoi fini :

> Garde la réponse gagnante d'Arena, mais ramène le mail à 150 mots. Garde le prix, la ligne pour se désabonner et la question finale. Mets le reste dans une liste à part, en dessous.

Voir pourquoi elle a gagné (non testée)

> Montre-moi les attaques que la réponse gagnante a encaissées pendant le tournoi, et ce qu'elle a changé pour chacune.

Lire le code avant d'installer (non testée)

> Clone https://github.com/Jakeschincariol/arena-skill dans un dossier temporaire. Lis SKILL.md et bracket.py, et dis-moi ce qu'ils écrivent, à quel endroit, et s'ils accèdent au réseau. Ne lance rien.

9. Les limites

Ce qu'Arena ne fait pas

- Elle ne fait pas travailler des modèles différents. Le README le dit : « 100 versions of Claude » veut dire 100 sous-agents du modèle de ta session. Ce qui change, c'est la carte.
- Ses juges sont Claude. Ils notent sur un barème écrit, mais chaque match a son juge, et les notes ne se comparent pas d'un match à l'autre : A015 a eu 86, 88,5 puis 91, et seulement 85 en finale. Le README prévient : ce que tu obtiens est « the strongest answer this tournament found, not a proof that it is right », la meilleure réponse que ce tournoi a trouvée, sans preuve qu'elle est juste.
- Elle ne répare pas une demande floue, et ses sous-agents ne voient que le fichier de tâche.
- La même graine ne redonne pas les mêmes réponses.
- Elle n'applique rien dans ton projet. C'est toi qui décides.

Ce que notre test ne prouve pas

- Une seule demande, un seul tournoi, en --quick. Le mode 100 n'a pas tourné chez nous.
- Le juge final est Claude lui aussi. Il a mis 5 sur 10 en clarté à la gagnante et 9 à la réponse normale : il a vu la longueur, mais il a quand même préféré la réponse longue.
- Personne n'a envoyé ces mails à de vrais prospects. On ne sait pas lequel obtient le plus de réponses.
- Notre mode restreint limitait le terminal à python3. Dans une session normale, Arena te demandera plus souvent ton accord.
- Une limite citée par le README, la variable CLAUDE_CODE_MAX_TOOL_USE_CONCURRENCY, n'apparaît pas dans la liste officielle des variables de Claude Code relue le 10 novembre 2026. La documentation cite en revanche une limite de 20 sous-agents actifs en même temps (CLAUDE_CODE_MAX_CONCURRENT_SUBAGENTS). Les vagues de 10 restent en dessous. Garde --wave à 10.

Sources

- Le dépôt Jakeschincariol/arena-skill de Jake Schincariol, licence MIT, commit unique df07b8e du 27 septembre 2026 : README.md, skills/arena/SKILL.md, bracket.py, rubric.md, strategies.json, lus avec l'API de GitHub le 21 octobre et le 10 novembre 2026 (1 406 étoiles ce jour-là).
- Notre test du 21 octobre 2026 : Claude Code 2.1.280, modèle Opus, mode restreint, skill chargée seule. Fichiers gardés : la demande, la réponse normale (385 mots), la réponse gagnante (3 950 mots), le fichier du tournoi, les 16 réponses de départ, les attaques, défenses et verdicts des 15 matchs, et le verdict final. Durée et coût lus dans le résultat de Claude Code.
- Documentation de Claude Code, relue le 10 novembre 2026 : modes de permission, variables d'environnement.

- CNIL : la prospection commerciale par courrier électronique, mise à jour le 10 juin 2026, relue le 10 novembre 2026.
- Cours de référence de la BCE du 9 novembre 2026 (1 € = 1,1193 $) : ecb.europa.eu.