

L'affirmation

Entre la démonstration du fournisseur et le budget de l'an prochain, une affirmation change d'espèce : capacité dans la démonstration, argent dans le budget. Quelles preuves exiger avant d'autoriser ce changement ?

Notre position est stricte : une affirmation de productivité n'est réelle qu'une fois traduite en ligne de compte de résultat (effectifs, temps de cycle, coût des erreurs), rapportée à une base nommée, avec un propriétaire et une date de revue. Démonstrations, bancs d'essai et études de cas ne sont pas cette preuve : ils montrent qu'une tâche peut être accomplie, pas qu'une ligne de coûts bougera ici. Un conseil qui budgète sur des preuves de capacité perd en précision avec un meilleur vocabulaire.

Les faits

Des preuves réelles existent, solides à l'échelle de la tâche : un déploiement échelonné sur 5 179 agents de support, une expérience randomisée sur 453 professionnels, une expérience contrôlée sur 95 développeurs. Les moyennes cachent l'essentiel : les gains se concentrent chez les débutants.

PREUVE

Les expériences mesurent des gains de tâche importants : 14% de demandes résolues par heure en plus chez 5 179 agents (34% chez les débutants); 40% de temps en moins et qualité notée 18% plus haut chez 453 professionnels; tâche de code achevée 55,8% plus vite chez 95 développeurs.

Source : Brynjolfsson, Li et Raymond, Quarterly Journal of Economics 140(2), 2025 ; Noy et Zhang, Science 381, 2023 ; Peng et al., arXiv:2302.06590, 2023 (document de travail) · Géographie : États-Unis et populations à distance · Méthode : déploiement échelonné et expériences randomisées · SEG-4 · Réserve : effets de tâche, populations propres, court horizon ; l'étude Copilot n'est pas évaluée par les pairs ; aucune ne mesure le compte de résultat d'une firme.

Les petites lignes sont l'essentiel : chaque étude mesure la production par unité de temps sur la tâche traitée, population précise, horizon court, et aucune ne mesure ce sur quoi un conseil budgète : masse salariale, temps de cycle, coût des erreurs, devenir des minutes gagnées.

Ces chiffres ne se collent pas dans un plan d'entreprise : la raison est arithmétique. L'analyse par tâches d'Acemoglu conclut à un gain de productivité globale des facteurs modeste sur dix ans. C'est la structure, non les intrants discutés, qui se transfère : le gain de la firme égale le gain de tâche multiplié par la part de travail de la tâche, puis par la part de coût de ce travail, chaque terme étant une fraction. Une économie de 40% sur une tâche qui occupe 5% du travail de 20% des effectifs n'est pas un programme de productivité ; c'est une erreur d'arrondi en costume.

PREUVE

Une analyse par tâches, agrégeant les économies mesurées sur la part de travail exposée, estime le gain total de productivité globale des facteurs sur dix ans entre 0,5 et 0,7%; chiffre modeste, peut-être surestimé si les tâches difficiles rapportent moins que les faciles.

Source : Acemoglu, « The Simple Macroeconomics of AI », Economic Policy 40(121), 2025 (NBER WP 32487, 2024) · Géographie : États-Unis, modèle d'économie entière · Méthode : modèle par tâches calibré sur l'expérimental · SEG-4 · Réserve : projection aux intrants discutés ; d'autres estimations sont plus élevées ; l'arithmétique se transfère, pas l'estimation ponctuelle.

Les statistiques officielles n'arbitreront pas non plus : la productivité du travail est un résidu sectoriel, publié en millésimes et révisé à l'arrivée des données sources. Elles répondent à une autre question, plus tard, pour une autre unité d'analyse, et n'attribueront aucun gain à votre déploiement.

Ce qu'aucune source ne fournit, car c'est un jugement de firme, c'est notre test budgétaire : avant d'entrer dans un plan, l'affirmation doit nommer sa ligne de compte de résultat, sa base de référence datée, son propriétaire et sa date de revue. Sans cela, c'est une humeur d'achat : elle se budgete comme un coût accompagné d'une expérience, pas comme une économie.

Fiches de preuve

CLM-GENAI-TASK-EXPERIMENTS · SEG-4. Les expériences mesurent des gains de tâche importants : 14 % de débit moyen chez 5 179 agents de support (34 % chez les débutants, gain minime chez les expérimentés) ; 40% de temps en moins et qualité notée 18 % plus haut chez 453 professionnels ; 55,8 % plus vite sur une tâche de code chez 95 développeurs. Contexte : Brynjolfsson, Li et Raymond, QJE 140(2), 2025 ; Noy et Zhang, Science 381, 2023 ; Peng et al., arXiv 2302.06590, 2023. Méthode : déploiement échelonné ; expériences randomisées. Preuves contraires : niveau tâche, court horizon, populations propres ; l'étude Copilot n'est pas évaluée. Confiance causale : expérimentale, dans le cadre étudié. Transférabilité : bornée aux tâches et populations comparables. Date de revue : 2026-08-02.

CLM-TASK-TO-FIRM-GAP · SEG-4. Agréger les économies de tâche mesurées sur la part de travail exposée donne des effets modestes, environ 0,5 à 0,7 % de productivité globale des facteurs sur dix ans, parce que le gain de la firme est le gain de tâche multiplié par la part de tâche et la part de coût. Contexte : Acemoglu, Economic Policy 40(121), 2025 ; NBER WP 32487. Méthode : modèle par tâches calibré sur l'expérimental. Preuves contraires : intrants discutés ; d'autres estimations sont plus élevées ; une projection n'est pas une mesure. Confiance causale : aucune revendiquée. Transférabilité : la structure se transfère, pas l'estimation ponctuelle. Date de revue : 2026-08-02.

CLM-PRODUCTIVITY-STATS-VINTAGE · SEG-5. Les statistiques officielles de productivité du travail sont des résidus sectoriels publiés en millésimes programmés et révisés (une estimation préliminaire et deux révisions au BLS) et ne peuvent attribuer un gain au déploiement d'une firme donnée. Contexte : documentation du programme de productivité du BLS. Méthode : méthodologie statistique officielle. Preuves contraires : aucune ; la limite est de conception. Confiance causale : aucune revendiquée, fait institutionnel. Transférabilité : structures analogues dans d'autres systèmes statistiques nationaux. Date de revue : 2026-08-02.

CLM-FIRM-PNL-TRANSLATION · SEG-1 (position de la firme). Une affirmation de productivité d'IA n'entre dans un budget qu'une fois traduite en ligne de compte de résultat nommée, contre une base de référence

datée, avec un propriétaire, une date de revue et un seuil d'arrêt ; démonstrations, bancs d'essai et études de cas sont des preuves de capacité, pas des preuves budgétaires. Contexte : pratique de la firme dans le corridor. Méthode : interprétation, pas mesure ; aucune statistique client citée. Preuves contraires : pour de petites expériences, l'instrumentation peut coûter plus qu'elles ; le test se proportionne aux montants en jeu. Confiance causale : aucune revendiquée. Transférabilité : bornée aux montants lourds et aux bases constructibles. Date de revue : 2026-08-02.

Les limites

Toute affirmation importée traverse le corridor avec la structure salariale de son étude : les expériences valorisent le temps gagné en salaires américains. Le même gain de débit dans un back-office d'Abidjan est un autre fait financier : l'heure gagnée se valorise dans une masse salariale plusieurs fois inférieure, tandis que l'outil se paie en dollars ou en euros et ne se réduit pas à proportion. Le point d'équilibre se déplace, parfois hors d'atteinte ; connectivité et énergie entrent dans des coûts que les firmes étudiées n'ont jamais mesurés. Plus dur encore : la base de référence n'existe souvent pas. Temps de cycle non mesurés, coût des erreurs non chiffré, débit connu d'instinct. Une société qui ne sait pas énoncer son coût par dossier traité ne peut pas savoir si l'IA l'a amélioré, mais elle peut être facturée chaque mois : les coûts arrivent libellés et datés, les bénéfices restent un récit tant que personne n'a construit la ligne.

Sources et limites

Sources et limites. Les résultats de tâche reposent sur Brynjolfsson, Li et Raymond (QJE, 2025), Noy et Zhang (Science, 2023) et Peng et al. (document de travail arXiv, 2023, non évalué) ; l'argument d'agrégation sur Acemoglu (Economic Policy, 2025), utilisé pour sa structure, pas pour sa précision ; la réserve statistique sur la documentation du BLS. Tout a été vérifié contre des sources officielles et académiques le 2026-08-02 ; aucune estimation de cabinet de conseil n'est citée. Le test de traduction est la position de la firme, graduée comme interprétation : pas un instrument validé, et aucun chiffre client publié parce qu'aucun n'a passé notre porte de publication. L'arithmétique salariale du corridor est un raisonnement, pas une donnée mesurée. Cette note ne vaut ni prévision des économies d'une firme, ni thèse selon laquelle les gains seraient illusoires : ils doivent seulement être traduits, rapportés à une base et dotés d'un propriétaire avant d'être budgétés.

Que faire lundi

- 1. Le directeur financier refuse toute ligne d'IA qui ne sait pas nommer, en une phrase, sa ligne de compte de résultat, sa base de référence datée et son propriétaire.** On la finance comme une expérience mesurée, ou pas.
- 2. Le commanditaire nomme la population d'étude à laquelle ressemble le cas d'usage et fait l'arithmétique salariale : le gain valorisé dans notre paie, l'outil dans la devise du fournisseur.** Les gains se concentrent chez les débutants ; les preuves citées ne parlent pas du jugement senior.
- 3. Le responsable métier fixe une date, un chiffre et un seuil d'arrêt avant le déploiement.** Ligne immobile à la revue : l'arrêt est la règle.
- 4. Le contrôleur de gestion rapproche l'affirmation brute de la ligne de coûts complète.** Abonnements, intégration, connectivité, temps de revue, reprises : un bénéfice suivi brut de ses coûts est une publicité.