

8,3 milliards d'humains simulés, zéro chat ivre

Le jour où Harvard et le MIT ont promis de construire Dwarf Fortress sans le savoir.

Par **Bruno DE SAN NICOLAS** · août 2026 · bdsn.eu



En 2015, des joueurs découvrent des chats morts couverts de vomi dans leurs tavernes virtuelles. Personne n'avait programmé ça : les chats traversaient des flaques de bière renversée, se léchaient les pattes pour faire leur toilette, et mouraient d'un excès d'alcool qu'une erreur de code surestimait. Onze ans plus tard, un consortium d'une centaine de chercheurs de Harvard, du MIT et de Stanford publie 8,3 milliards d'humains simulés. Pas un seul n'est capable de surprendre son créateur. Cette différence a un nom. Et elle vaut une méthode.

Acte I : mille ans avant votre premier coup de pioche

Avant de raconter les hommes, il faut raconter le monde. Lancez Dwarf Fortress et le jeu ne vous demande pas de choisir un niveau : il vous demande de créer une planète. Sous vos yeux, la géologie se sculpte, les montagnes se soulèvent, les rivières creusent patiemment leur lit, les pluies et les vents dessinent des climats. Puis, avant même que vous ne posiez le premier pion, la machine simule un

millénaire d'histoire. Des civilisations naissent, prospèrent, se déchirent. Des héros accomplissent des quêtes dont chaque bataille a ses causes. Des artefacts sont forgés, perdus, retrouvés. Des dynasties s'éteignent. Rien de tout cela n'est écrit à l'avance : tout est calculé, événement après événement, et chaque guerre, chaque roi, chaque objet légendaire reçoit un nom et une biographie générés par le déroulement même de cette histoire.

Un joueur raconte avoir découvert, en fondant sa forteresse, qu'un elfe régnait sur ses nains. Absurde ? Le journal du monde expliquait tout : une guerre ancienne où cet elfe avait sauvé la citadelle, sa lente conversion à la culture naine, l'extinction de la lignée royale, et un vote des survivants. Personne n'avait écrit cette histoire. Elle avait eu lieu, dans la machine, des siècles simulés avant la première partie.

Descendez maintenant à l'échelle de l'individu, car c'est là que le vertige commence. Chaque nain, et il peut y en avoir des dizaines de milliers, porte une cinquantaine de traits de personnalité, une trentaine de valeurs, plus d'une centaine d'émotions distinctes, des préférences alimentaires forgées dans son enfance simulée, et des souvenirs qui le hantent. Son corps n'est pas une barre de points de vie : c'est une anatomie. Muscles, os, organes, nerfs, vaisseaux. Une morsure à la main gauche sectionne des tendons précis ; le jeu calcule la douleur, le choc, l'hémorragie ; une infection négligée devient gangrène. Les carnets de conception de Tarn Adams, le créateur, notent que la température doit affecter la viscosité du sang. Le jeu génère aujourd'hui 269 types de roches et minéraux, 784 créatures, et compose des poèmes selon des formes poétiques qu'il invente lui-même, culture par culture.

Qui a bâti cette cathédrale ? Deux frères de l'État de Washington. Zach, l'aîné, historien de formation, passé par les entrepôts et les chantiers navals, qui transforme en systèmes ce qu'il a appris de Rome et de Byzance : les successions, les coups d'État, les schismes. Tarn, le cadet, docteur en mathématiques de Stanford dans une spécialité si abstraite qu'aucune université proche de chez lui ne sut l'employer. Exilé dans un post-doctorat texan, isolé, rongé par la dépression, il démissionne en 2006, rentre chez ses parents, cloue une planche sur la fenêtre de sa chambre d'enfant et code chaque nuit, de 23 heures à 7 heures. Le 6 août 2006, les deux frères mettent en ligne, gratuitement, ce jeu au graphisme de terminal des années 80. Vingt ans plus tard, il est entré au musée d'art moderne de New York, a inspiré Minecraft, RimWorld et Prison Architect, et son créateur estime l'œuvre achevée à 47 %.

Mais la vraie preuve de la profondeur d'une simulation, ce sont ses accidents. Les chats ivres de 2015, que personne n'avait programmés : trois systèmes honnêtes, le déplacement, la toilette, l'ingestion, dont l'interaction produisit une hécatombe féline dans les tavernes. Boatmurdered, cette partie collective de 2006 où un bug dotant les éléphants d'une mémoire infinie engendra des décennies de guerre transgénérationnelle entre pachydermes rancuniers et nains assiégés, jusqu'à la folie collective et l'effondrement ; la communauté en fit des fresques, des chansons, une mythologie. Les fermes à os de sirène, où des joueurs très mercenaires exploitèrent une valeur marchande mal calibrée jusqu'à contraindre le créateur à intervenir. Une phrase de Zach Adams résume la doctrine qui rend tout cela possible : ils ne créent pas des histoires, ils créent les conditions pour que les histoires naissent d'elles-mêmes. L'émergence ne se force pas. Elle s'invite.

Acte II : le consortium

Août 2026. Une centaine d'auteurs, trois universités d'élite en première ligne, un comité consultatif où siègent quelques-uns des noms les plus respectés de l'IA, des soutiens nommés Anthropic, OpenAI, Microsoft, Amazon et Meta. Le papier s'intitule « MatrAIX: Simulating the World with 8.3 Billion Persona Agents ». Le problème visé est réel et sérieux : les études utilisateurs humaines coûtent cher et prennent des semaines ; les tests automatiques mesurent la réussite d'une tâche sans jamais modéliser la diversité de ceux qui l'exécutent. Remplacer les cobayes par des populations simulées est une idée parfaitement légitime.

Et l'infrastructure est soignée, disons-le sans ironie. Chaque humain simulé est décrit par 1 290 caractéristiques. Un graphe de dépendances échantillonne les profils en respectant les corrélations du monde réel : le niveau d'études selon l'âge, la maîtrise de l'anglais selon la langue maternelle et la région, avec des règles excluant les combinaisons impossibles. Une partie du corpus s'ancre dans des données humaines observées, des biographies aux grandes enquêtes sociologiques. C'est de l'ingénierie de population de bonne facture.

Mais regardons les ordres de grandeur, car ils racontent autre chose. Le titre annonce 8,3 milliards de personas, la population terrestre, clin d'œil assumé. Le jeu de données public en contient un million. La bibliothèque revendique 1 010 tâches d'évaluation ; huit ont été exécutées pour le papier. Et le chiffre rond s'obtient en supprimant délibérément 1,35 milliard d'enregistrements excédentaires : cette humanité simulée est un objectif de communication avant d'être une mesure.

Acte III : la fracture

Le cœur de l'affaire tient en une image. Un nain de Dwarf Fortress est un journal intime : chaque page s'écrit avec les événements, et la page d'hier détermine celle d'aujourd'hui. Un persona MatrAIX est un curriculum vitae : 1 290 cases remplies une fois pour toutes, puis tendues à un modèle de langage prié de « jouer » ce profil. Le premier vit ses attributs. Le second les récite.

Les propres chiffres du papier, pour qui descend dans ses annexes, mesurent cette fracture avec une précision cruelle. Sur des cohortes strictement identiques de 1 000 clients simulés, la part choisissant l'abonnement payant d'un même produit passe de 23,2 % à 93,9 % selon le modèle d'IA qui incarne les personas. Mêmes profils, même question, conclusions commerciales opposées : l'instrument mesure le modèle, pas la population. L'accord entre modèles jouant les mêmes personas, calculé sur 88 champs comparables, donne un kappa de Cohen médian de zéro, c'est-à-dire l'absence totale de concordance au-delà du hasard. Et le détail le plus parlant : interrogés sur leur propre tranche d'âge, les personas répondent juste dans 16,3 % des cas sous un modèle, 16,9 % sous un autre, quand le pur hasard donnerait 16,7 %. Deux des trois modèles testés jouent donc des humains qui ignorent leur âge ; le troisième répond parfaitement, ce qui prouve seulement qu'il lit mieux sa fiche. Un nain de 2006, lui, connaît son âge, ses cicatrices, et le nom de l'éléphant qui a tué son grand-père.

Le papier revendique pourtant 91,5 % d'« adhérence comportementale ». Regardez ce qu'on mesure : usage d'emojis, politesse, verbosité, style de commentaires de code. On vérifie que le persona parle comme son étiquette, pas qu'il décide comme un humain ; et le juge de ce score est le même modèle que celui qui joue les agents, sans qu'aucune évaluation humaine n'entre dans le chiffre-titre. Surtout, l'adhérence est l'exact contraire de l'émergence : un persona y est récompensé quand il ne surprend personne, quand un nain entre dans la légende en surprenant son propre créateur. Et lorsque l'étiquette contredit le tempérament du modèle sous-jacent, l'étiquette perd : sommés d'être impolis, les agents refusent cinq fois sur cinq, leur dressage de politesse écrasant le rôle. Le costume ne tient pas sur le mannequin.

Une trouvaille du papier mérite pourtant d'être sauvée, car elle pointe la sortie. Dans l'étude la plus coûteuse, des personas pilotent la vraie application Apple News sur un simulateur d'iPhone. Le taux d'abonnement est statistiquement inexploitable. Mais les justifications en texte libre sont remarquables : 23 raisons sur 24 invoquent des spécificités du profil, pas assez de presse lusophone pour l'un, rien en agronomie hispanophone pour l'autre, catalogue trop occidental pour un troisième. Quand on laisse ces agents raconter au lieu de cocher des cases, quelque chose affleure. Les frères Adams approuveraient : la valeur était dans l'histoire, pas dans le taux.

Acte IV : la marche à suivre

Je ne suis pas commentateur. Une comparaison ne vaut que si elle produit une conduite. En voici une, en six règles, pour quiconque envisage d'utiliser, d'acheter ou de bâtir des populations simulées, en études de marché comme en aide à la décision.

Un. Exigez la surprise. En démonstration fournisseur, une seule question : « montrez-moi le registre des comportements que vos concepteurs n'avaient pas prévus, et reproduisez-en un devant moi ». Inscrivez-la au cahier des charges comme critère de recette. Un éditeur de simulation sérieuse tient ce registre, c'est son équivalent des chats ivres ; celui qui n'en a pas vous vend un questionnaire déguisé.

Deux. Lisez le texte libre, méfiez-vous des taux. Clause de livraison : chaque étude inclut l'export intégral des verbatims et des trajectoires d'interaction, essai par essai, pas seulement le tableau de bord. Et une règle de gouvernance interne : aucune décision produite ne s'appuie sur un taux simulé seul ; les pourcentages ouvrent des hypothèses, les verbatims les instruisent, les humains les tranchent.

Trois. Croisez les socles. Clause contractuelle : toute étude est exécutée sous au moins deux modèles d'IA de familles distinctes, et l'écart entre modèles est publié avec chaque résultat. Fixez le seuil à l'avance : si cet écart dépasse l'effet que vous cherchez à mesurer, l'étude est déclarée non conclusive, point. Rappel utile en négociation : dans le papier du MIT lui-même, ce même écart atteint 70 points sur une simple question de prix. Et jamais le même modèle des deux côtés du miroir, jouant le client et le produit testé : il reconnaît sa propre prose, la préfère, et masque les frictions que l'étude devait révéler.

Quatre. Testez la mémoire et l'état ; bâtisseurs, inversez la dépense. Protocole de recette en trois questions, à dérouler en séance face au fournisseur : demandez son âge au persona ; demandez-lui ce qu'il vous a dit lors de la session précédente ; contrariez un de ses intérêts déclarés et observez si son comportement ultérieur en garde trace. Trois échecs, ce n'est pas un être simulé, c'est une fiche. Pour ceux qui construisent, la conséquence est architecturale : cessez d'allonger le CV, commencez le journal intime. Trois chantiers, dans cet ordre : un état interne qui évolue (fatigue, confiance, agacement montent et descendent au fil des interactions, comme l'humeur d'un nain), une mémoire d'événements qui contraint le comportement du lendemain, y compris contre le profil initial, et des règles d'interaction entre ces états, même simples. Commencez avec trente variables qui s'influencent avant de rêver à mille colonnes : mille attributs figés produisent un annuaire, trente variables couplées produisent une société.

Cinq. Validez contre le réel avant toute décision conséquente. Règle budgétaire simple : sur tout budget d'étude par simulation, sanctuarisez un cinquième pour la contre-épreuve humaine de l'hypothèse principale, sur le même questionnaire et la même tâche. La simulation travaille en amont, pour cribler large et vite ; l'humain travaille en aval, pour décider. Les auteurs de MatrAix l'écrivent eux-mêmes dans une annexe d'une honnêteté remarquable ; c'est l'écart entre cette annexe et leur titre qui fait débat, pas leur rigueur.

Six. Installez une veille sur le jeu vidéo au rang de votre veille académique. Concrètement : nommez un référent, envoyez-le une fois l'an à la Game Developers Conference de San Francisco, dont le sommet consacré à l'IA précède de plusieurs années les conférences académiques sur ces sujets, et faites circuler les retours d'expérience techniques que les studios publient après chaque sortie. Coût : celui d'un billet d'avion, pas d'un programme. Pour les organisations de défense, un pas de plus : mettez votre cellule wargaming en contact direct avec les studios de simulation, car le Kriegsspiel prussien dont descend tout notre wargaming était déjà un jeu que l'institution avait eu l'intelligence d'adopter. Le jeu a toujours été un laboratoire du sérieux ; Dwarf Fortress est à la simulation de sociétés ce que le Kriegsspiel fut à la manœuvre : une avance prise en jouant.

. . .

Dans leurs travaux futurs, les auteurs de MatrAix écrivent qu'il faudra développer des personas dynamiques dotés de mémoire à long terme. Autrement dit : construire Dwarf Fortress. Ils l'ont écrit eux-mêmes, sans paraître mesurer l'aveu. Un dernier clin d'œil, que la statistique offre gracieusement : leur schéma compte cinquante dimensions de loisirs et quatre-vingt-une de médias ; parmi leurs 8,3 milliards d'humains simulés figurent donc, nécessairement, quelques milliers de joueurs de Dwarf Fortress. Incapables, hélas, de se souvenir de leur dernière forteresse. D'un côté, un consortium qui titre « Simuler le monde » sur huit tâches exécutées. De l'autre, un artisan qui, après vingt ans, estime sa cathédrale achevée à 47 % et dont le jeu porte pour devise officieuse « Losing is fun » : perdre est amusant, car l'effondrement instructif vaut mieux que la validation confortable. Entre les deux, toute la distance qui sépare l'ambition de l'humilité. Dans la simulation comme ailleurs, c'est l'humilité qui tient la distance.

Sources

1. X. Li, Y. Hao et al., « MatrAIX: Simulating the World with 8.3 Billion Persona Agents », arXiv:2608.04205, 4 août 2026. Corps de l'article et annexes E à M (résultats produit, fidélité des personas, adhérence, limites).
2. K. Stuart, « Unexpected nudity and vomit-covered cats: how Dwarf Fortress tells some of gaming's most bizarre stories », The Guardian, novembre 2022.
3. Noclip, série documentaire « Dwarf Fortress », épisode 1, 2024-2025. Entretiens avec Tarn et Zach Adams.
4. « Dwarf Fortress », vidéo documentaire francophone, youtube.com/watch?v=Iq1uCqH62JQ.
5. Wikipedia, articles « Dwarf Fortress » et « Tarn Adams » : dates de développement, sortie Steam 2022, entrée au MoMA.
6. Forums Something Awful, récit collectif « Boatmurdered », novembre 2006.