

Introduction

De Netflix à la NASA, quand les organisations se laissent « emporter par la foule »

Les modèles d'affaires des organisations publiques et privées connaissent, depuis le développement du Web participatif, de profondes transformations. La notion de Web 2.0 permet de saisir ce tournant. Elle désigne un Web devenu collaboratif, interactif et participatif, dans lequel les internautes ne sont plus de simples récepteurs de contenus, mais participent activement à leur production, leur circulation et leur valorisation (O'Reilly, 2007). C'est sur ce terrain que le *crowdsourcing* se développe à partir des années 2000, en proposant aux organisations d'externaliser certaines tâches vers une foule d'internautes. Héritier de cette dynamique participative, mais désormais inscrit dans l'économie des plateformes, des données et de l'Intelligence Artificielle (IA), le *crowdsourcing* connaît aujourd'hui une mutation profonde. La foule n'est plus seulement mobilisée pour créer, résoudre, évaluer ou financer, elle contribue aussi, directement ou indirectement, à entraîner, corriger et orienter les algorithmes d'IA (Encadré 1).

ENCADRÉ 1. De la « sagesse des foules » au « travail du clic » au service de l'IA

Alors que le *crowdsourcing* était initialement célébré pour sa capacité à innover (Netflix Prize, InnoCentive, etc.), une partie massive de l'activité s'est déplacée vers des tâches invisibles mais cruciales : l'entraînement des IA. Derrière la fluidité d'un outil comme ChatGPT ou la créativité de Midjourney, se cache une multitude de *clickworkers* qui annotent des images, classent des textes ou corrigent du code. La performance des IA génératives dépend conséquemment de multiples opérations humaines d'annotation,

d'évaluation et de correction des contenus. Comme l'avaient pressenti Gray et Suri (2019) avec le concept de « *Ghost Work* », une partie croissante du *crowdsourcing* repose sur un travail humain invisible indispensable au fonctionnement des plateformes numériques.

Modalité d'externalisation vers une foule en ligne, les champs d'application du *crowdsourcing* sont infinis et parfois surprenants : trouver de l'or (Wilde & Kreuzer, 2016), rechercher des criminels (Graham & Mehmood, 2014), mener des enquêtes géopolitiques en source ouverte comme le fait le collectif Bellingcat (Higgins, 2021), déceler les traces d'un avion disparu¹, entraîner une IA à converser poliment, définir l'anatomie génitale féminine idéale², choisir le nom d'une marque ou bien encore d'un animal de compagnie... Développée ci-après, une sélection de cas témoigne de la force et de la diversité du phénomène.

- « Rassembler les passionnés des photos de la guerre civile et leur donner accès à des technologies de pointe et à des ressources personnalisées pour soutenir leurs recherches » tel est l'objet de la plateforme Civil War Photo Sleuth (CWPS). Cette dernière repose d'une part, sur la technologie de reconnaissance faciale de Microsoft Azure³ et d'autre part, sur le *crowdsourcing*. Lorsqu'un utilisateur télécharge une photo, le logiciel de reconnaissance tente de détecter le visage du soldat. Il s'agit d'analyser des dizaines de points de référence qualifiés d'empreintes faciales et de comparer ces points aux milliers de photos présentes dans les archives de la plateforme. La technologie de reconnaissance faciale permet de trouver des correspondances dans les visages même si la pilosité diffère ou si la photo est prise sous un angle différent. Perfectible, cette technologie s'enrichit par le recours à la foule *via* le *crowdsourcing*. Il s'agit ainsi de compléter les résultats générés par l'ordinateur par des jugements humains. *Via* le *crowdsourcing*, les internautes répondent ainsi à la problématique du « dernier kilomètre » dans l'identification des personnes (Mohanty *et al.*, 2018). La force du système de vision humaine permet par conséquent d'enrichir le système d'IA dans une tâche aussi complexe que la reconnaissance d'images. Cette initiative préfigure ce que l'on nomme le *Reinforcement Learning from Human Feedback*⁴, où le jugement humain est indispensable pour affiner,

1 *Via* la plateforme Tomnod, des milliers de personnes avaient tenté de retrouver les traces du Boeing 777 de la Malaysia Airlines disparu en juillet 2014.

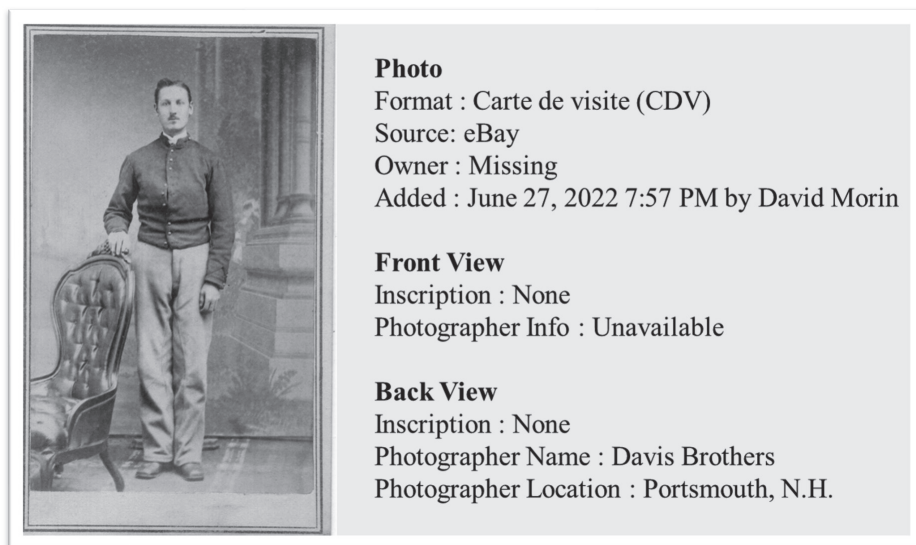
2 L'objectif est d'aider les chirurgiens à optimiser les résultats de la chirurgie esthétique de la vulve (Frojo *et al.*, 2022).

3 Microsoft Azure est une plateforme applicative en nuage (cloud) proposant plusieurs services à ses utilisateurs.

4 L'expression anglaise *Reinforcement Learning from Human Feedback* (RLHF) désigne l'apprentissage par renforcement à partir de rétroaction humaine.

corriger et valider les résultats bruts des modèles d'IA les plus avancés. Afin d'illustrer nos propos, la figure 1 met en perspective la photo d'un homme dont l'identité est recherchée par la communauté CWPS.

Figure 1. Civil War Photo Sleuth : (en)quête d'identités



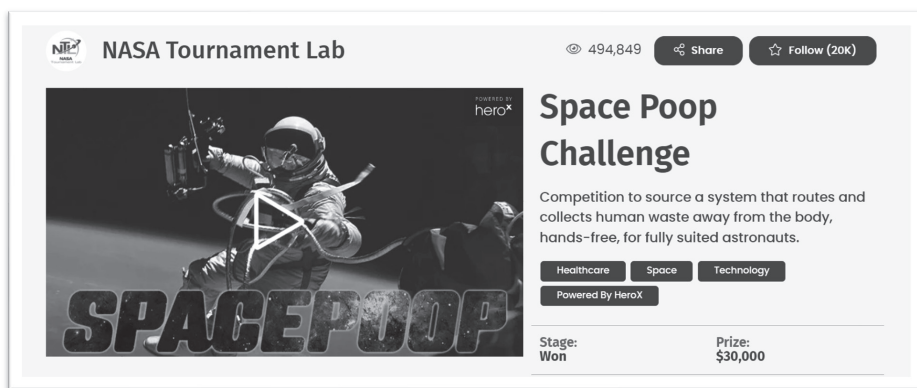
Source : civilwarphotosleuth.com

- Netflix a souhaité en 2006 tirer profit des potentialités offertes par le *crowdsourcing* pour améliorer la technologie Cinematch ayant pour objet de prédire les films qu'un abonné pourrait aimer. Alors que les listes de recommandations étaient relativement satisfaisantes, il s'agissait d'exploiter plus encore les potentialités de l'outil. Avec pour objectif de développer un algorithme plus performant, « le Netflix prize » offrait la coquette somme d'un million de dollars à quiconque améliorerait de 10 % le système de recommandation de films (Bell *et al.*, 2010). La valeur du prix a sans doute beaucoup participé à sa popularité et a, par ailleurs, permis d'attirer des experts de la modélisation mathématique et de la programmation informatique (Rosen, 2011). Afin que les participants disposent d'informations similaires à celles détenues par ses employés, Netflix a partagé en ligne des données réelles anonymisées. Attirant des milliers de candidats du monde entier, le concours est devenu un modèle de collaboration et d'innovation à l'ère d'Internet. Comme l'indiquent Bell *et al.* (2010), des équipes se sont constituées avec des membres issus de différents pays communiquant par courriel et partageant le fruit de leur travail avec des personnes qu'ils n'avaient jamais rencontrées physiquement. C'est finalement un groupe

d'individus disposant de compétences complémentaires qui a emporté le challenge (Périn & Burger-Helmchen, 2011).

- Alors que la *National Aeronautics and Space Administration*, plus connue sous son acronyme NASA, est reconnue pour sa main-d'œuvre aussi talentueuse que qualifiée, elle a, à plusieurs occasions, eu recours à la sagacité des foules en ligne. Malgré des freins internes liés au syndrome *Not Invented Here*⁵ (NIH), la NASA a notamment fait appel en 2016 à la sagesse de la foule en lançant sur la plateforme HeroX le « Space Poop Challenge » (Figure 2). Le défi des internautes : solutionner l'une des problématiques incontournables de la condition humaine : le caca ! Plus précisément, il s'agissait d'améliorer les méthodes de gestion des déchets humains (matières fécales, urine et menstruations) lors des activités spatiales de longue durée. Charge donc aux innovateurs du monde entier de proposer un système de traitement des déchets produits par les astronautes. Le caractère amusant du défi et la dotation de 30 000 dollars ont participé à sa viralité. La NASA a obtenu des réponses intéressantes lui permettant de développer de futures combinaisons spatiales (Matzler, 2020).

Figure 2. Le Space Poop Challenge



Source : herox.com

- La pandémie de la Covid-19 a suscité la mise en place de dispositifs de *crowdsourcing* visant à développer des thérapies, fabriquer des masques ou bien encore faire évoluer les respirateurs afin de prendre en charge deux patients à la fois. Dans un ouvrage dédié au déploiement du *crowdsourcing* pendant la crise épidémique, Bueno-Muñoz *et al.* (2022)

5 Le syndrome NIH (*Not Invented Here*) désigne la réticence d'une organisation à reconnaître, adopter ou valoriser des idées et solutions issues de l'extérieur, au motif implicite qu'elles n'ont pas été conçues en interne.

mettent notamment en avant l'application *COVID Symptom Study*. Créée en mars 2020, l'application collecte et répertorie auprès des volontaires de précieuses informations. Les données recueillies sont utilisées pour étudier les symptômes et la vitesse de propagation du virus, ainsi que pour identifier les zones à haut risque de contagion. L'étude des données ainsi produites par la foule a donné lieu à de nombreux travaux scientifiques publiés dans des revues majeures (Bueno-Muñoz *et al.*, 2022). En réponse à l'urgence épidémique internationale, c'est aussi une branche spécifique du *crowdsourcing*, le *crowdfunding*, qui a été mobilisé pour acheter des masques, accompagner les personnes isolées, soutenir financièrement des professionnels contraints d'arrêter leur activité (Chekhar & Renault, 2020). Cet épisode historique a démontré la capacité de résilience et de réactivité de la foule face à une crise mondiale d'envergure.

De nombreux domaines sont susceptibles d'être concernés par le *crowdsourcing* : cartographie, médecine, marketing, musique, littérature ou encore photographie. Cette diversité d'applications invite à interroger la notion même de foule, souvent prise dans une tension entre folie et sagesse. Depuis les travaux fondateurs sur les foules et les publics (Le Bon, 1895 ; Park, 1972 ; etc.), la réflexion oscille en effet entre crainte des phénomènes d'imitation, d'irrationalité ou de manipulation et célébration de formes d'intelligence collective. La foule n'est pourtant ni seulement une masse irrationnelle, emportée par la contagion des affects, ni nécessairement une intelligence collective spontanément vertueuse. Elle peut produire du désordre, du conformisme ou de l'erreur, mais elle peut aussi, lorsqu'elle est orientée vers une finalité commune, devenir une ressource précieuse à l'action collective. Dans ce contexte, le *crowdsourcing* apparaît comme une forme contemporaine de mobilisation organisée de la participation collective. Loin des perspectives mettant exclusivement en avant la folie des foules (Freud, 1921), mais sans pour autant céder à une vision naïve de leur sagesse (Surowiecki, 2008), le *crowdsourcing* fait de la foule un partenaire de choix dans la réponse à de multiples défis : de la réalisation de tâches du quotidien à la résolution d'épineuses problématiques. À l'instar de M. Jourdain qui faisait de la prose sans le savoir, chaque internaute s'est déjà prêté à une activité de *crowdsourcing*, parfois consciemment, parfois sans nécessairement percevoir que ses usages numériques ordinaires contribuent à une dynamique collective organisée. Compte tenu de la place croissante du *crowdsourcing* dans la société, l'objectif de cet ouvrage est de développer une compréhension globale de ses contours, modalités et enjeux.