

L'espace francophone de l'intelligence artificielle. Le travail humain au cœur

du processus de production des données

Le titre de mon intervention aux Doctoriales : « La France avec les robots micro travail de la donnée en start-up nation », était un clin d'œil à un ouvrage du philosophe Georges Bernanos, La France contre les robots dans lequel il critiquait le libéralisme et les formes d'automatisation qu'il observe à son époque. Cette thématique rejoint mon sujet de thèse puisque je travaille sur toutes les activités inhérentes au développement de l'intelligence artificielle notamment de productions de données ou de vérification du fonctionnement d'un algorithme. C'est ce que l'on appelle communément le micro travail, qui correspond à des tâches courtes et faiblement rémunérées effectuées par des travailleurs.

La définition a priori très complexe, de l'Intelligence Artificielle, peut être résumée assez facilement en utilisant celle du récent white paper de la Commission Européenne « Il s'agit d'un ensemble de technologies combinant des données, des algorithmes et de la puissance de calcul. » Les données renvoient à de grandes bases de données sur lequel se base l'entraînement d'algorithmes. Les algorithmes renvoient aux programmes des concepteurs et la puissance de calcul aux data centers.

Derrière cette définition aux promesses ou aux risques de l'IA puisque sa finalité est de reproduire la cognition humaine, de reproduire le fonctionnement de l'être humain qui se traduit par certains exemples en science-fiction comme Terminator. Dans le cadre de ma thèse, l'IA est envisagé avec un angle commercial. Par exemple, cette fonction d'une borne de cantine permettant de scanner les plateaux repas en faisant de la reconnaissance d'image.

Mais le cœur de ma thèse est de regarder comment des travailleurs font fonctionner l'IA. Un premier exemple de ce que peuvent faire des travailleurs de l'IA sont ces personnes que j'ai rencontrées qui avaient pour mission de tracer un cercle autour d'une tomate sans savoir pourquoi. Un second exemple de micro travail et des dysfonctionnements de l'IA pourrait être cette borne de scan expresse de plateau repas à la cantine de Télécom Paris. Elle est censée être expresse mais ne fonctionne pas toujours efficacement ce qui oblige la caissière à intervenir. Elixor spécialisée dans la restauration collective soustrait l'utilisation de cette borne à une start-up. Cette dernière est une filiale d'une application de gestion de calories. En synthèse, l'application indique le nombre de calorie d'un plat à partir d'une photographie. Lors d'une enquête que nous avons réalisée précédemment, nous avons réalisé que cette start-up s'appuyait des données récoltées sur une plate-forme française de travail en ligne pour lesquels des gens devaient entourer des tomates. Pour la réalisation de cette borne express, l'entreprise a utilisée des données récoltées par des micro-travailleurs afin d'entraîner son algorithme de reconnaissance visuelle de plateaux repas. La chaîne de valeur en vue du remplacement d'une caissière est démentielle : Un premier prestataire en charge de la restauration collective, un second fait fonctionner la borne et l'algorithme de plateau repas, une troisième filiale recueille des données sur lesquelles s'entraîne l'algorithme via une plate-forme de micro travail et enfin des micro-travailleurs.

Cette enquête sur le micro travail en France nous avait permis d'identifier les rôles des travailleurs de l'IA. Nous avons identifié l'entraînement de l'IA, l'imitation de l'IA et la vérification de l'IA. L'entraînement peut être par exemple de générer des données, enregistrer sa voix pour entraîner un assistant local. L'imitation de l'IA peut être un assistant virtuel dont l'objectif est de prendre des

rendez-vous à votre place. Mais en fait, c'est un être humain qui va regarder votre agenda et qui vous propose des rendez-vous. La vérification de l'IA, sera réalisée par des gens écoutant des extraits audio et vérifiant si l'algorithme a bien compris l'extrait audio...

La difficulté sera pour moi de théoriser le micro travail après la contribution théorique d'Antonio Casilli relatée dans son ouvrage En attendant les robots. Je me suis axée sur trois concepts : la plate-formisation, l'externalisation et l'internationalisation qui se rejoignent et se recoupent.

En matière de plate-formisation, la digitalisation est une forme de digital labor. Le digital labor est soit du travail rémunéré via des plates-formes de travail en ligne souvent sans protection sociale, soit des activités numériques quotidiennes permettant d'entraîner des algorithmes. Les Recaptcha, par exemple permettent de produire des données et de vérifier que l'utilisateur est bien un être humain. La gig economy, autre manière de parler de digital labor est liée à la plate-formisation et concerne l'économie des petits jobs. Les travailleurs ubérisés et certains free lances opérant sur des plateformes de travail en ligne appartiennent à la gig economy. Mais se focaliser uniquement sur l'aspect plate-forme limite l'analyse que l'on peut avoir de cette pratique car le micro travail peut aussi concerner des salariés d'entreprises. Si on considère la plate-formisation, j'ai essayé de résumer des controverses.

Une première controverse est soulevée par A Casilli concerne le travail non rémunéré que beaucoup d'utilisateurs d'internet effectue sans réaliser qu'ils travaillent et génèrent et fournissent des données aux plates-formes. Je n'aborderai pas ce point dans ma thèse car je ne traiterai que du travail rémunéré.

La seconde controverse porte sur la marchandisation du travail et remise en question de l'institution salariale. Qui pose la question quel travail et quel statut pour les travailleurs des plates-formes ? Cette controverse ne sera pas non plus au cœur de ma thèse, car je ne veux pas me limiter à ne travailler sur les travailleurs des plates-formes.

La seconde manière de caractériser le micro travail est de se référer aux activités d'externalisation de service informatique. Le micro travail a existé sous une autre forme, dans les années 1980-1990, sous le nom de « data entry ». Ainsi, des coupons de billets d'avion étaient envoyés en Jamaïque puis renvoyés sous forme numérisés aux Etats Unis. Le micro travail peut être aussi une forme de télé travail. Les recherches de Penny Gurstein montrent qu'à l'époque des entreprises étaient réticentes au télé travail pour de la production de données car elles avaient peur d'une part pour des enjeux de vie privée et d'autre part que les télé travailleuses concernées ne soient ni forcément mobilisables sur toute une journée ni activable de manière très rapide parce que retenues par des activités domestiques. Cette activité peut aussi avoir lieu dans des call center.

L'internationalisation est liée encore une fois à l'externalisation et la plate-formisation. Dans une de ses publications, Mark Graham observe que l'offre et la demande de travail d'une plate-forme de travail en ligne qui ne fait pas uniquement du micro travail, mais également des tâches plus intensives en terme de compétences comme par exemple du développement web. La littérature montre que la logique d'externalisation s'opère entre pays culturellement proches d'un point de vue linguistique. Les relations de micro travail entre les Etats Unis, l'Inde et les Philippines montre qu'une relation nord sud s'opère. En ancrant mon sujet dans la start-up nation et la station F par exemple me mène à l'hypothèse que le micro-travail des entreprises françaises est réalisé dans des pays

africaines francophones et dans certains pays de l'est. J'ai mené un entretien avec l'ancien responsable d'une entreprise malgache de sous-traitance qui travaille pour une start-up française de reconnaissance d'image. Son équipe traitait des images satellites, en récupérant les panneaux de circulation ou des images de tous les véhicules d'une certaine couleur de photos satellite à la demande de l'entreprise française. Elle traitait également des images de repas, ce qui nous renvoie à la start-up de reconnaissance visuelle de plateaux repas précédemment évoquée qui réalise le même travail. Au lieu d'avoir des données récoltées via une plate-forme de micro travail, ils font appel, cette fois-ci, à une entreprise de sous-traitance appelée Business process outsourcing. Cette dernière contractualise avec un entrepreneur du pays qui va ensuite recruter des travailleurs et ensuite travailler à distance avec cette entreprise.

Deux controverses résultent de la situation décrite plus haute. Quels sont les déterminants de la localisation de la chaîne de valeur des données qui alimentent l'IA ? Comment identifier des pays qui seraient en mesure d'accueillir de l'externalisation ? Je vais probablement m'éloigner car ces controverses se posent plus en terme de positionnement dans le champ de la production des idées.

Comment sont produites les données nécessaires au fonctionnement de l'IA dans les pays francophones ? Cela amène plusieurs hypothèses de recherche

- La chaîne de valeur de l'IA met en relation des entreprises françaises et des soustraitants des pays francophones du continent africains. Les pays de l'Est peuvent aussi être une piste intéressante.
- Les entreprises s'appuient sur différents types d'intermédiaires pour accéder à cette force de travail. L'enjeu est de voir comment ces liens culturels sont formalisés et par conséquent créent des sous-traitance entre des entreprises francophones et des entreprises malgaches.
- Les entreprises mettent en place des indicateurs pour s'assurer de la qualité des données. Cette question sera davantage en lien avec le biais algorithmique. Est-ce que le micro travail crée aussi des biais algorithmiques au même titre que d'autres données numériques. Plus la chaîne de sous-traitance est longue plus les risques de biais sont possibles.