



IA générative et recherche sociologique

Paola Tubaro, 15 avril 2025

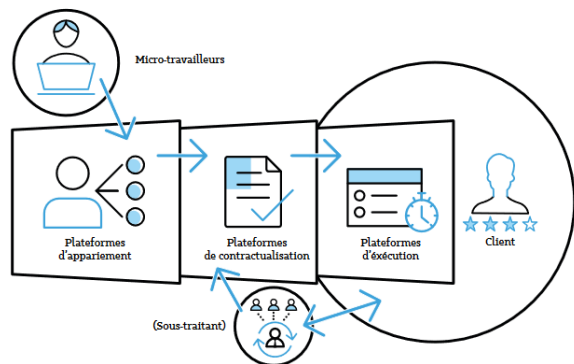


INSTITUT
POLYTECHNIQUE
DE PARIS

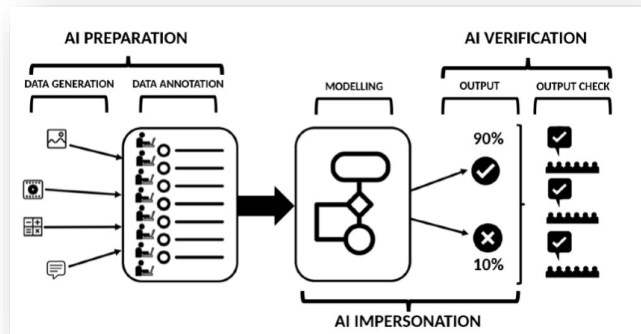


Groupe des écoles
nationales d'économie
et de statistique

Mes principaux sujets de recherche



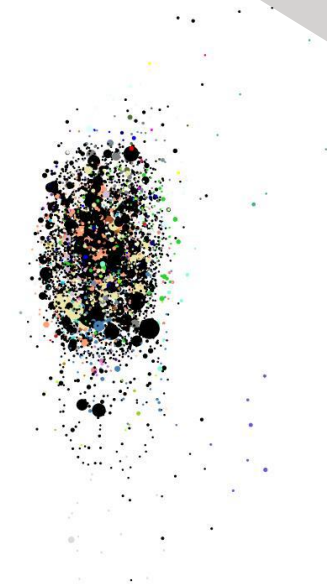
La platformisation de l'économie et du travail



Le rôle de l'humain dans la production d'IA



Les chaînes de production mondialisées de l'IA



Les dynamiques sociales de la désinformation en ligne

Les visages multiples de l'enquête sociologique

https://tria.limequery.org/576546

LimeSurvey

7%

Actividades por Internet

Para comenzar vamos a hablar de las actividades que lleva a cabo por Internet.

*¿Cuánto tiempo hace que utiliza la página Clickworker?

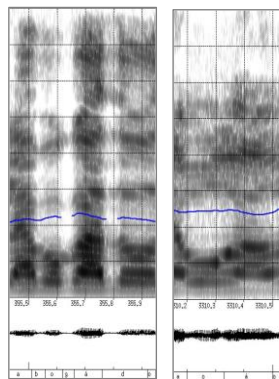
● Seleccione una de las siguientes opciones:

- ☐ Menos de un mes
- ☐ 1-3 meses
- ☐ 4-7 meses
- ☐ 8-12 meses
- ☐ 1-2 años
- ☐ 2-3 años
- ☐ 3 años o más

*¿Cómo se enteró de la existencia de la página Clickworker?

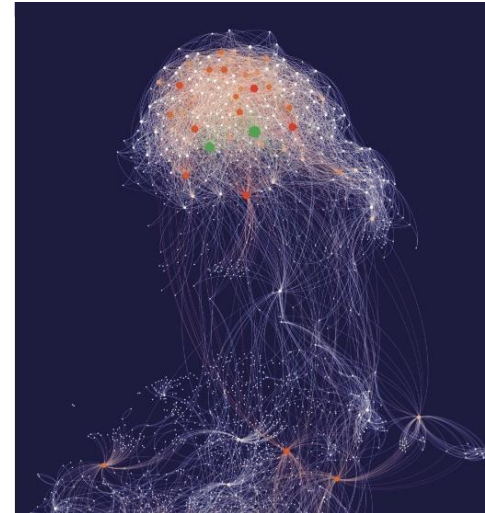
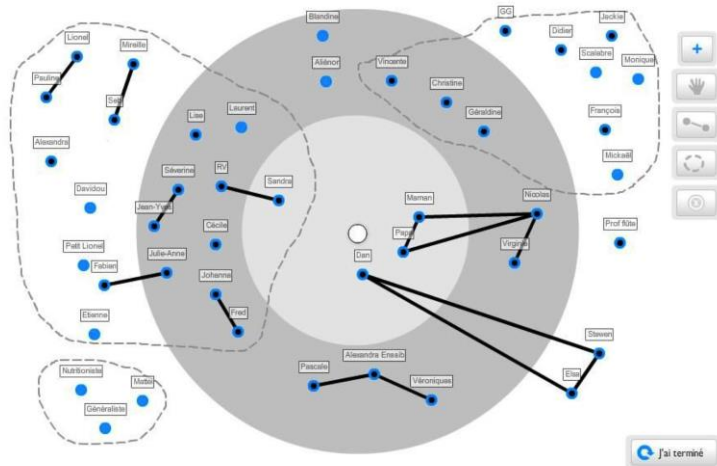
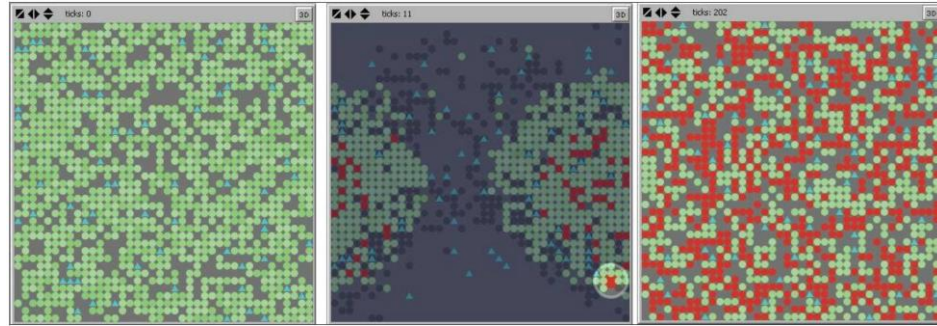
● Seleccione una de las siguientes opciones:

- ☐ Un amigo/una amiga o conocido cercano
- ☐ Un colega o ex colega

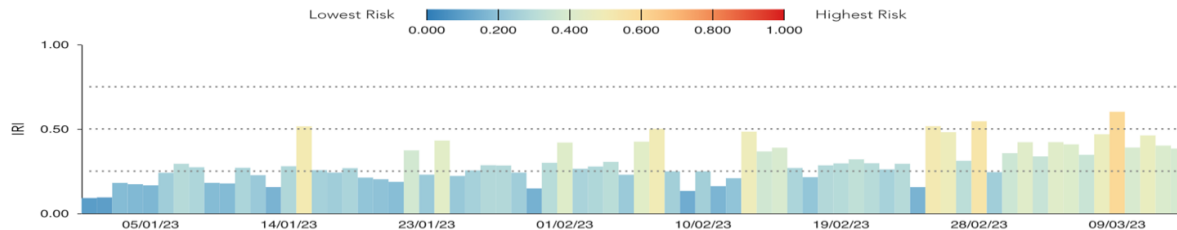
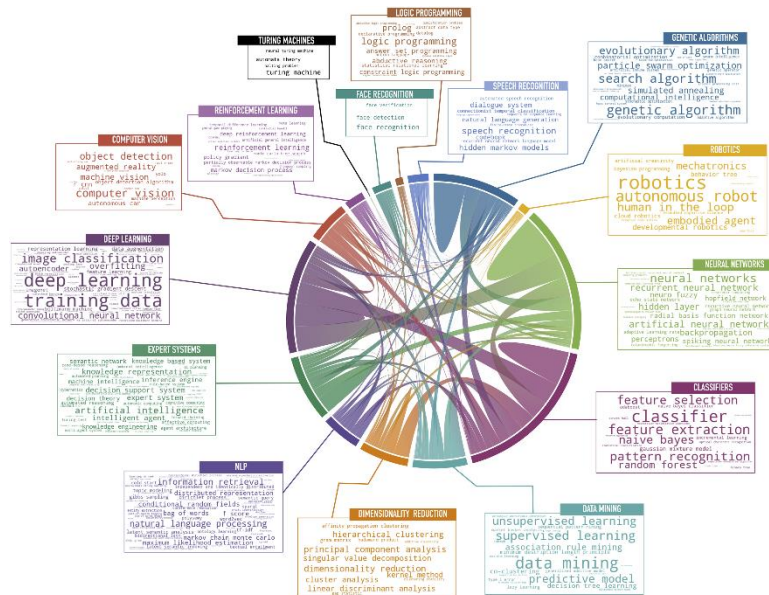
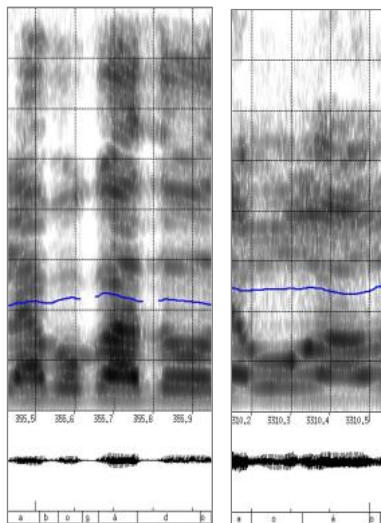


1	Exporter M.49	Exporter ISO3	Importer M.49	Importer ISO3	Resource	Year	Value (1000USD)	Weight (1000kg)
11312	56	BEL	348	HUN	Quartz and	2022	0.16480966	0.10994799
11313	56	BEL	348	HUN	Sand and gr	2022	235.380073	635.589204
11314	56	BEL	348	HUN	Stone	2022	751.7515935	1287.014208
11315	56	BEL	352	ISL	Cement	2022	123.19686	332.762
11316	56	BEL	352	ISL	Clays	2022	4.5677183	13.3412784
11317	56	BEL	352	ISL	Other mine	2022	12.85610041	22.72103066
11318	56	BEL	352	ISL	Plasters	2022	0.881705	0.106
11319	56	BEL	352	ISL	Sand and g	2022	10.53292	15.2779181
11320	56	BEL	352	ISL	Slag and as	2022	59.027134	318.19168
11321	56	BEL	352	ISL	Stone	2022	6.3127965	6.0127789
11322	56	BEL	360	IDN	Clays	2022	25.1402688	23.6232355
11323	56	BEL	360	IDN	Lime	2022	36.6890218	98.453714
11324	56	BEL	360	IDN	Other mine	2022	229.746602	77.914935
11325	56	BEL	360	IDN	Quartz and	2022	55.347	88.13
11326	56	BEL	360	IDN	Sand and g	2022	238.771597	369.12834
11327	56	BEL	360	IDN	Stone	2022	0.482	0.646
11328	56	BEL	364	IRN	Other mine	2022	1.244703	0.8
11329	56	BEL	364	IRN	Sand and g	2022	2.111205	7.50034
11330	56	BEL	368	IRQ	Other mine	2022	5.831499	4.91523
11331	56	BEL	368	IRQ	Sand and g	2022	0.116636	0.45
11332	56	BEL	368	IRQ	Stone	2022	0.561064	0.9
11333	56	BEL	372	IRL	Cement	2022	187.75067	57.71998
11334	56	BEL	372	IRL	Clays	2022	8.0469914	21.9124914
11335	56	BEL	372	IRL	Lime	2022	18.469146	26.1793
11336	56	BEL	372	IRL	Other mine	2022	1603.216368	3854.134657
11337	56	BEL	372	IRL	Plasters	2022	0.621299	0.115888
11338	56	BEL	372	IRL	Quartz and	2022	0.24050686	0.027298113
11339	56	BEL	372	IRL	Sand and g	2022	282.6879772	820.8289055
11340	56	BEL	372	IRL	Slag and as	2022	258.062691	1578.618636
11341	56	BEL	372	IRL	Stone	2022	162.3149343	219.478474
11342	56	BEL	376	ISR	Cement	2022	0.9997519	13.390031
11343	56	BEL	376	ISR	Clays	2022	47.501902	84.824064
11344	56	BEL	376	ISR	Other mine	2022	1434.257123	14422.23951

Du numérique en sociologie...



... aux sciences sociales computationnelles



Que peut-on faire de plus avec l'IA générative?

Recherche naissante sur les usages possibles de l'IA générative:

- Annoter des données / extraire automatiquement de l'information de grandes bases de données (textuelles)
- Améliorer les simulations (notamment ABM) en enrichissant les profils des agents et en permettant des interactions plus complexes
- Prétester des questionnaires avant de les lancer

Cependant

- Biais très répandu, mais dont les caractéristiques sont souvent mal connues (entraînement non-transparent des modèles, objectifs commerciaux...)
 - Certains groupes sociaux sont mal représentés
 - Les modèles d'IA Gen ne peuvent pas remplacer une passation de questionnaires « classique » auprès d'humains
- Pour leur nature même, les modèles d'IA Gen peuvent produire de l'information incorrecte qui peut être difficile à détecter
- Réplicabilité : on connaît encore mal les effets de variations même légères des prompts + l'évolution des modèles peut rendre les résultats instables
- Manque de transparence concernant de nombreux modèles propriétaires - erreur toujours possible si l'on connaît mal l'outil

Bilan : des conclusions encore (très) provisoires

- Une recherche récente, susceptible d'évoluer à mesure que les outils s'améliorent et la compréhension de leur fonctionnement s'affine
- Une recherche jusqu'ici surtout méthodologique, qui n'a pas (encore) produit des résultats substantifs marquants
- Diffusion jusqu'ici limitée
- Une opportunité qui fait resurgir de questions sociologiques / méthodologiques classiques, par exemple concernant les objectifs et l'utilité des simulations ABM
- D'autres enjeux sont d'ordre éthique (effets sur les participants à une étude, coût humain de l'entraînement de modèles, effets sur l'environnement)



MERCI



Center for Research in Economics and Statistics
5 avenue Henry Le Chatelier
91120 Palaiseau

