



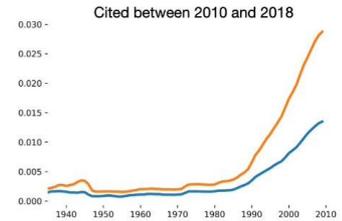
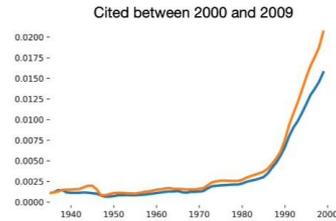
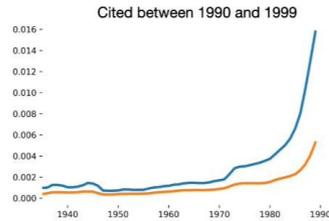
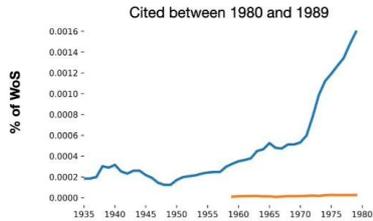
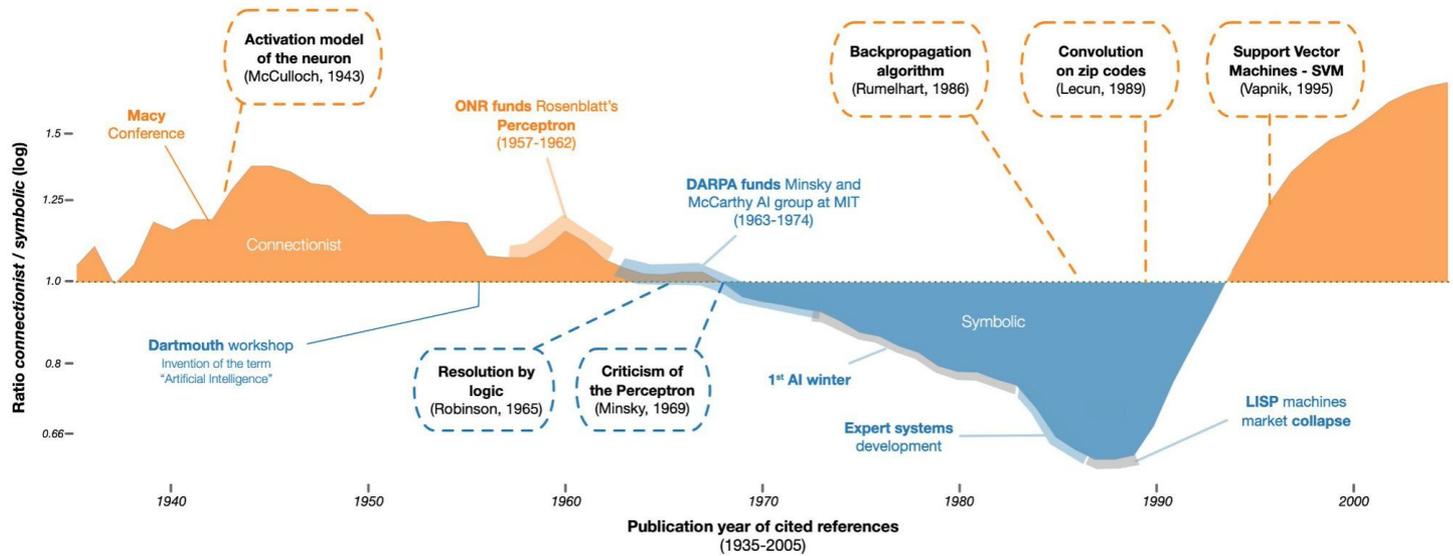
Pauline Gourlet

médialab de Sciences Po

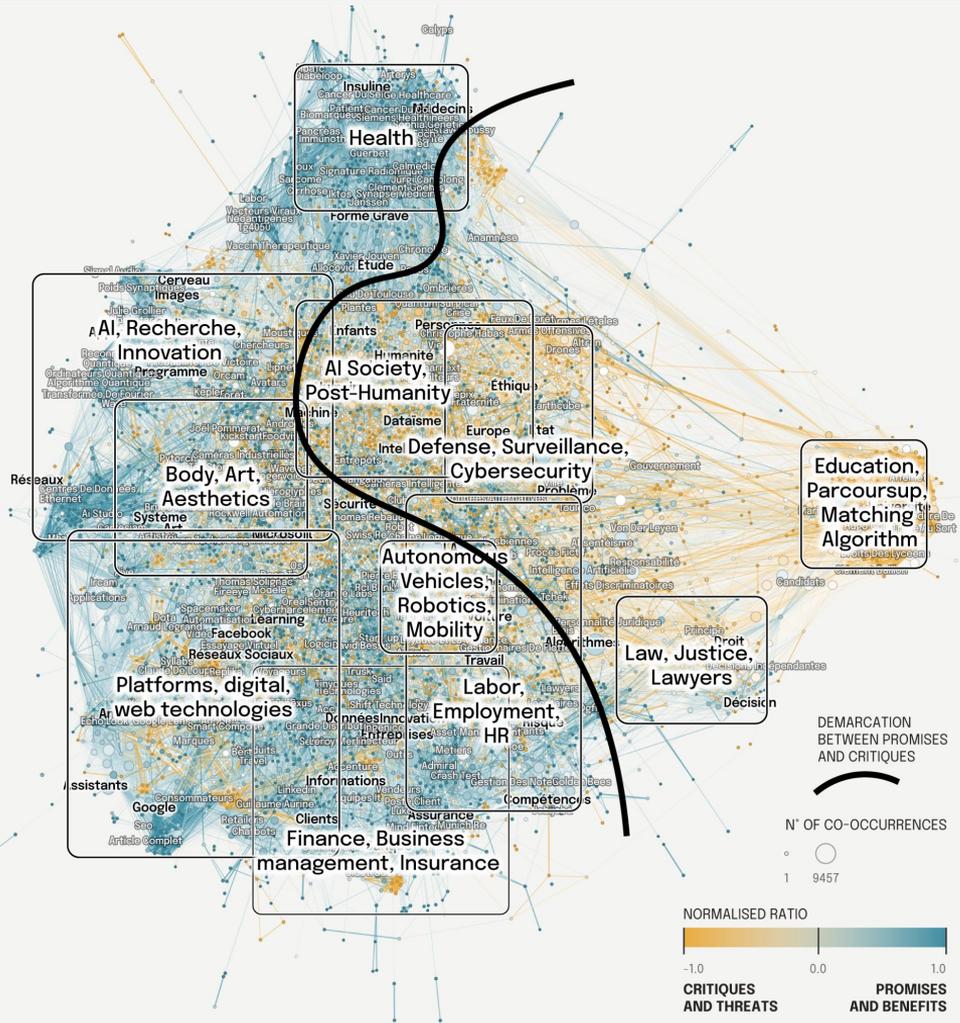
Que peut l'IA pour l'action publique ?

05 avril 2024

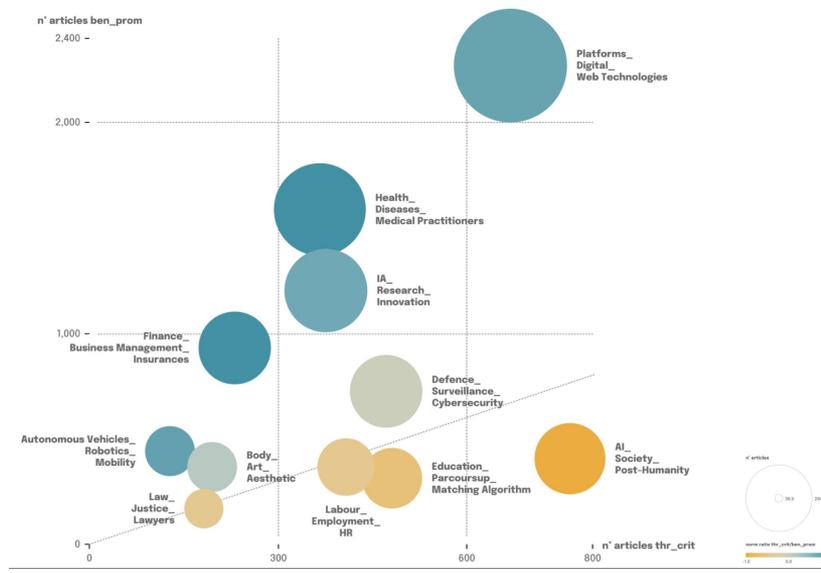
27e Région - séminaire

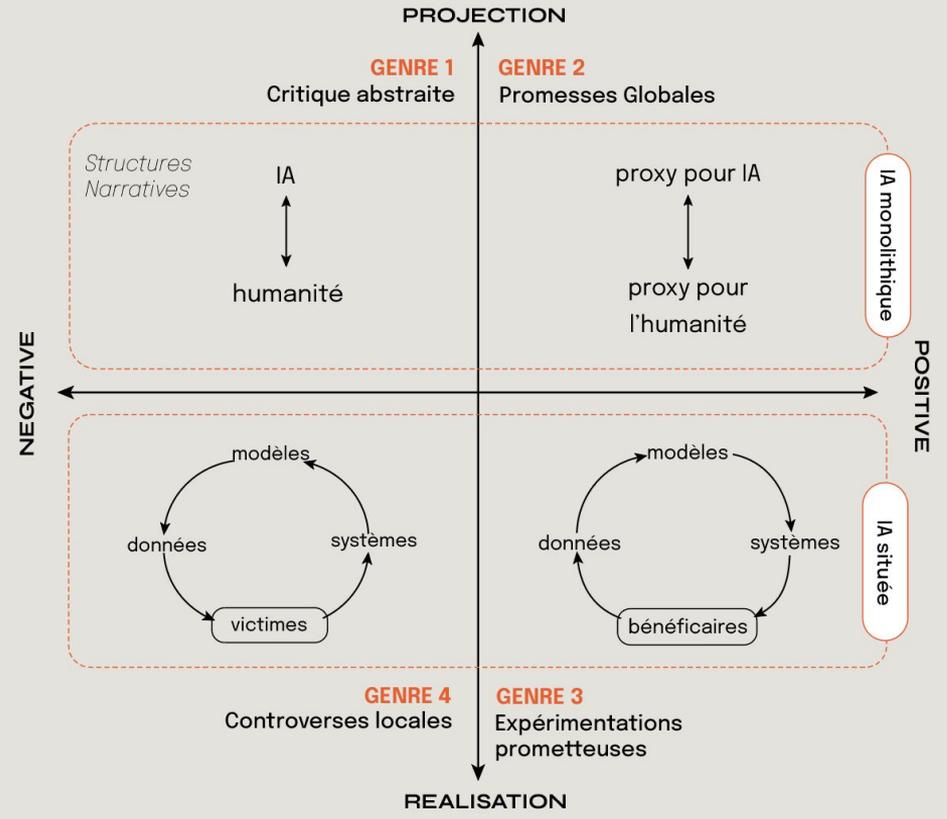


Dominique Cardon, Jean-Philippe Cointet, Antoine Mazières, La revanche des neurones
 L'invention des machines inductives et la controverse de l'intelligence artificielle
 dans Réseaux 2018/5 (n° 211), pages 173 à 220

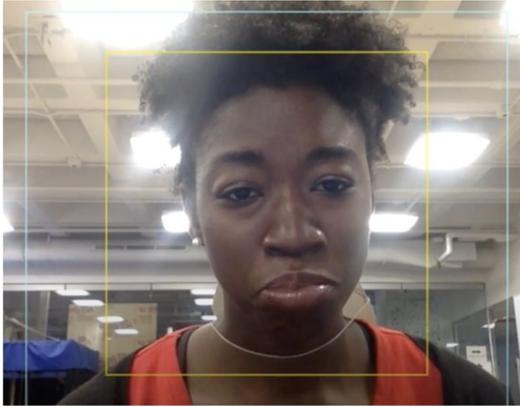


Gourlet et al. (2023) PRISES, MÉ.PRISES, EM.PRISES, DÉ.PRISES, RE.PRISES...
 Les formes de participation au développement de l'IA en France
<https://medialab.github.io/ShapingAI/>





Gourlet et al. (2023) PRISES, MÉ.PRISES, EM.PRISES, DÉ.PRISES, RE.PRISES...
 Les formes de participation au développement de l'IA en France
<https://medialab.github.io/ShapingAI/>



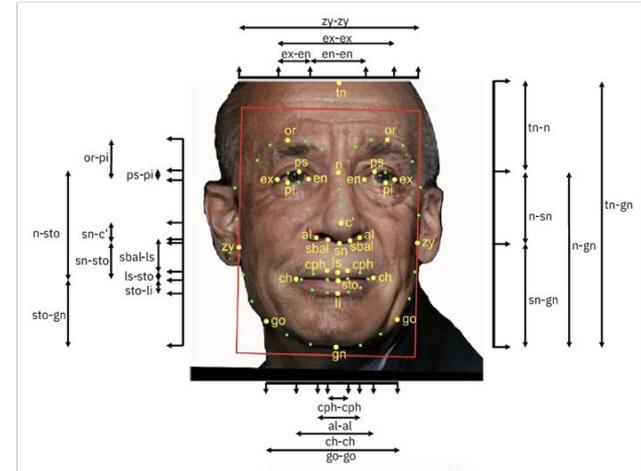
Joy Buolamniwi
MIT / the Algorithmic Justice League

IBM Let's Create Products & Solutions Consulting & Services Learn & Support Explore more

IBM Research Blog Topics Labs

AI

IBM Research Releases 'Diversity in Faces' Dataset to Advance Study of Fairness in Facial Recognition Systems



Excavating AI

The Politics of Images in Machine Learning Training Sets

By Kate Crawford and Trevor Paglen

You open up a database of pictures used to train artificial intelligence systems. At first, things seem straightforward. You're met with thousands of images: apples and oranges, birds, dogs, horses, mountains, clouds, houses, and street signs. But as you probe further into the dataset, people begin to appear: cheerleaders, scuba divers, welders, Boy Scouts, fire walkers, and flower girls. Things get strange: A photograph of a woman smiling in a bikini is labeled a "slattern, slut, slovenly woman, trollop." A young man drinking beer is categorized as an "alcoholic, alky, dipsomaniac, boozer, lush, soaker, souze." A child wearing sunglasses is classified as a "failure, loser, non-starter, unsuccessful person." You're looking at the "person" category in a dataset called ImageNet, one of the most widely used training sets for machine learning.

Something is wrong with this picture.

Where did these images come from? Why were the people in the photos labeled this way? What sorts of politics are at work when pictures are paired with labels, and what are the implications when they are used to train technical systems?

In short, how did we get here?

Check for updates

BIG DATA & SOCIETY

Original Research Article

On the genealogy of machine learning datasets: A critical history of ImageNet

Emily Denton¹, Alex Hanna¹, Razvan Amironesei², Andrew Smart¹ and Hilary Nicole¹

Big Data & Society
July–December 1–14
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journalsPermissions
DOI: 10.1177/2039570120395953
journals.sagepub.com/home/bds
SAGE

Abstract

In response to growing concerns of bias, discrimination, and unfairness perpetuated by algorithmic systems, the datasets used to train and evaluate machine learning models have come under increased scrutiny. Many of these examinations have focused on the contents of machine learning datasets, finding glaring underrepresentation of minoritized groups. In contrast, relatively little work has been done to examine the norms, values, and assumptions embedded in these datasets. In this work, we conceptualize machine learning datasets as a type of informational infrastructure, and motivate a genealogy as method in examining the histories and modes of constitution at play in their creation. We present a critical history of ImageNet as an exemplar, utilizing critical discourse analysis of major texts around ImageNet's creation and impact. We find that assumptions around ImageNet and other large computer vision datasets more generally rely on three themes: the aggregation and accumulation of more data, the computational construction of meaning, and making certain types of data labor invisible. By tracing the discourses that surround this influential benchmark, we contribute to the ongoing development of the standards and norms around data development in machine learning and artificial intelligence research.

Keywords

Machine learning, artificial intelligence, big data, AI ethics, algorithmic fairness, genealogy

Introduction

Machine learning (ML) systems are rapidly being developed and deployed in a variety of socially consequential domains. Yet, there is a growing abundance of examples of how these systems are failing people of color (Noble, 2018; Benjamin, 2019), women (Bolukbasi et al., 2016), LGBT+ communities (Scheuerman et al., 2019), people with disabilities (Hutchinson et al., 2020; Trewin, 2018), and the working class and those in poverty (Fabunmi, 2018). Many of these failures have been the direct result of underrepresentation, misrepresentation, or the complete lack of representation of these groups in the data upon which these systems are built (Paullada et al., 2020).

In response to these failures of algorithmic systems, a proliferation of algorithmic fairness interventions have emerged in recent years that hinge on balancing representation of different demographic groups within training datasets—the data used for algorithms to “learn” associations (Shankar et al., 2017; Merler et al., 2019; Yang et al., 2020). While interventions of this sort play a non-trivial role in achieving recently advanced technical definitions of algorithmic fairness (e.g. Hardt et al., 2016), failures of data-driven systems are not located exclusively at the

level of those who are represented or under-represented in the dataset. For example, deficiencies are often tied to unstated assumptions underlying the dataset or the schemas used to encode particular types of harmful classifications (e.g. Onofoa, 2016; Denton et al., 2020; Scheuerman et al., 2020). Solutions oriented around balancing representation of sociodemographic groups within ML datasets often reflect a focus on “fairness” at the wrong level of abstraction (Schst et al., 2019). Worse, data collection efforts aimed at increasing the representation of marginalized groups within training data are often executed through exploitative or extractive mechanisms (Chuiet, 2018; Soles, 2019).

In contrast to the significant efforts that have focused on statistical properties of training datasets, comparatively little attention has been paid to the various modes of their

¹Google Research, NY, USA

²Center for Applied Data Ethics, University of San Francisco, CA, USA

Corresponding author:

Emily Denton, Google Research, 1111 8th Ave, New York, NY 10011, USA.

Email: dentone@google.com

 Creative Commons CC BY-ND. This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits any use, reproduction and distribution of the work as published without adaptation or alteration, provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Denton, E., Hanna, A., Amironesei, R., Smart, A., & Nicole, H. (2021). *On the genealogy of machine learning datasets: A critical history of ImageNet*. *Big Data & Society*, 8(2).



Pierre Arlais
@PierreArlais

Suivre

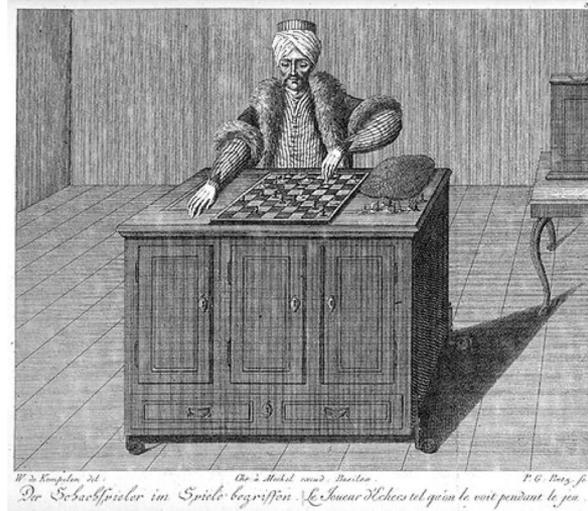
Amazon a annoncé la fin de ses magasins sans caisse. La technologie "Just Walk Out" présentée comme automatique, nécessitait en fait la surveillance par plus de 1 000 employés indiens des acheteurs un par un pour établir les factures !



[gizmodo.com](https://www.gizmodo.com)

Amazon Reportedly Ditches 'Just Walk Out' Checkouts at Its Grocery Stores

21:42 · 02 avr. 24 · 44,9K Vues



work hard. have fun. make history.



Artifice and Intelligence



Center on Privacy & Technology · [Follow](#)

Published in Center on Privacy & Technology at Georgetown Law · 9 min read · Mar 8, 2022



214



7



Words matter.

Starting today, the Privacy Center will stop using the terms “artificial intelligence,” “AI,” and “machine learning” in our work to expose and mitigate the harms of digital technologies in the lives of individuals and communities.



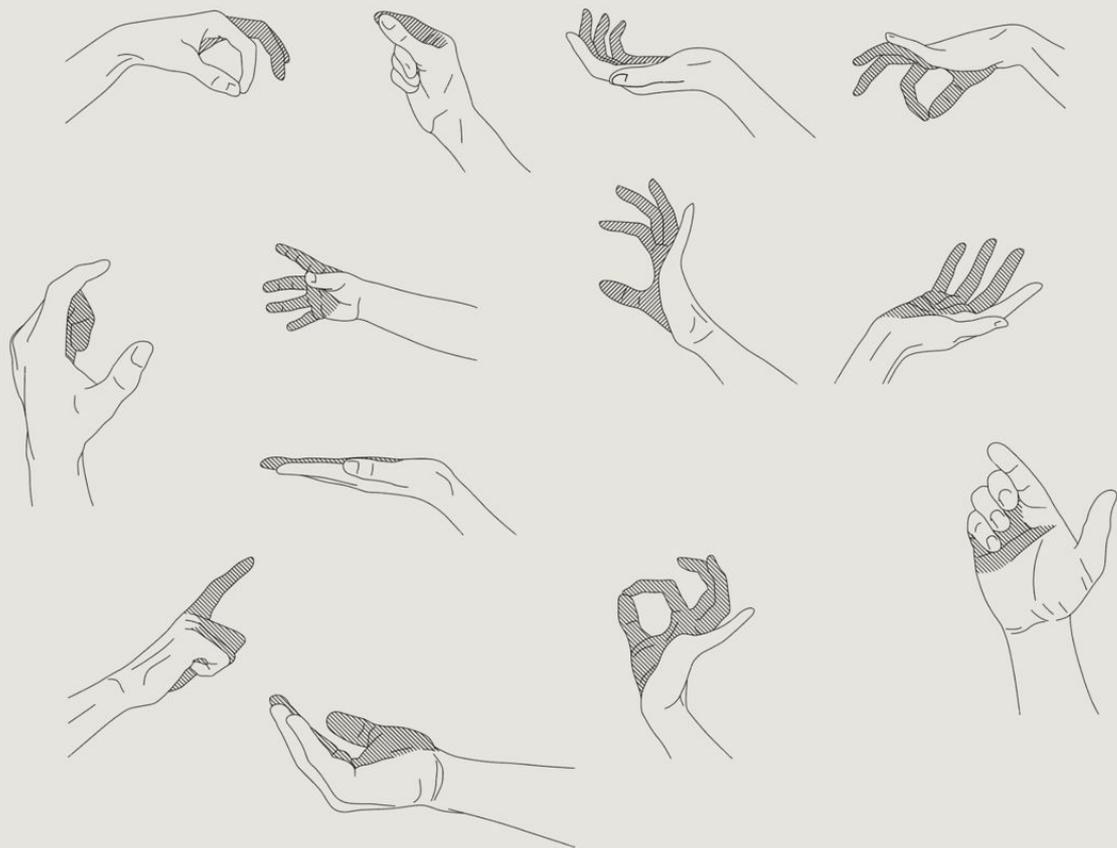
EU AI Act

Proposal for a

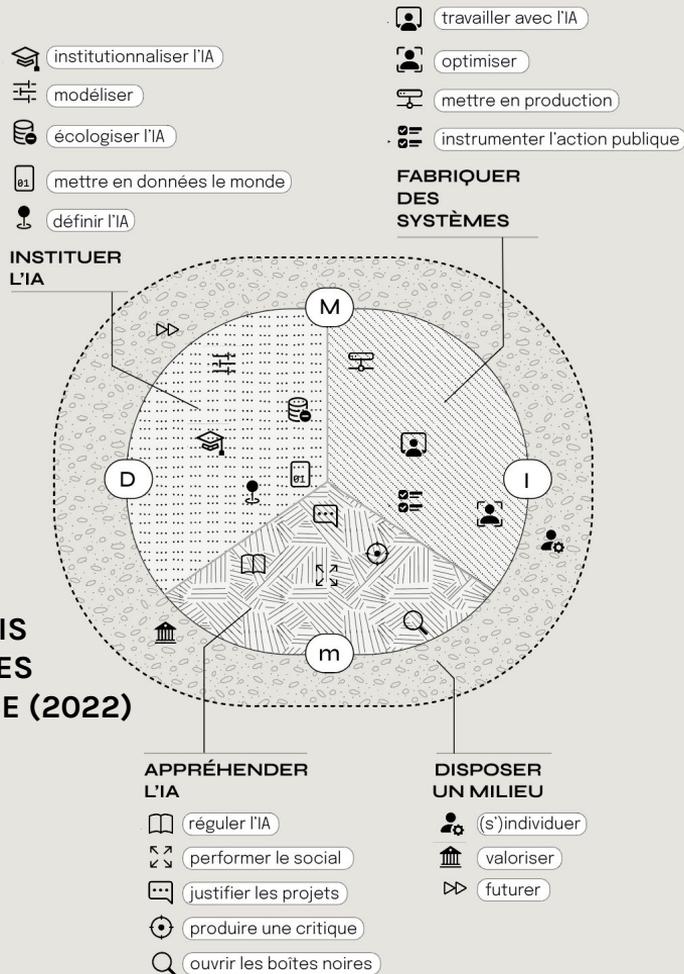
Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts

2021/0106 (COD)

European
Commission

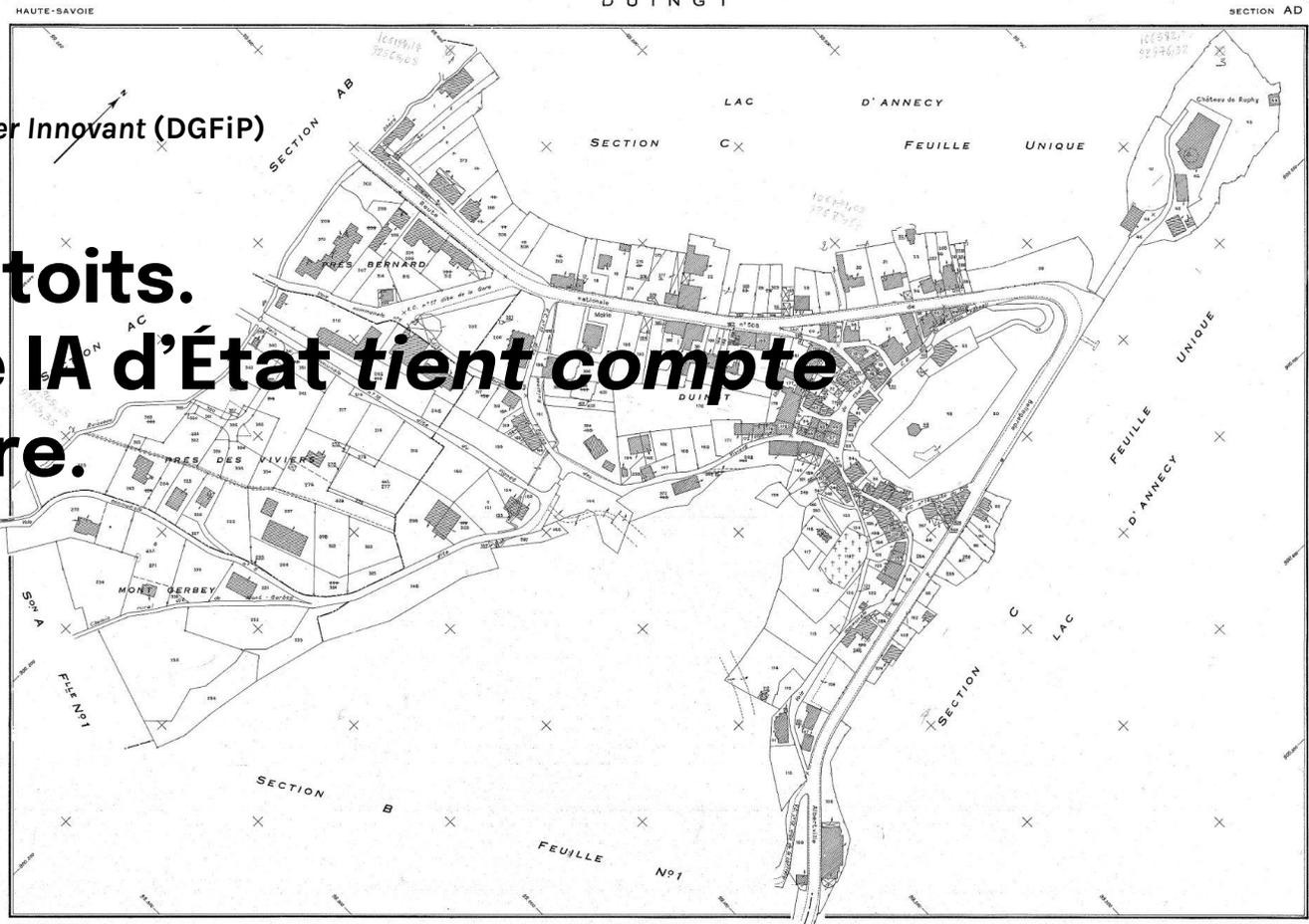


CARTE DES SOUCIS DES PRATICIEN.NES DE L'IA EN FRANCE (2022)



L'ensemble des soucis avec leur vidéo et documents sont accessibles ici : <https://medialab.github.io/ShapingAI/>

DUINGT

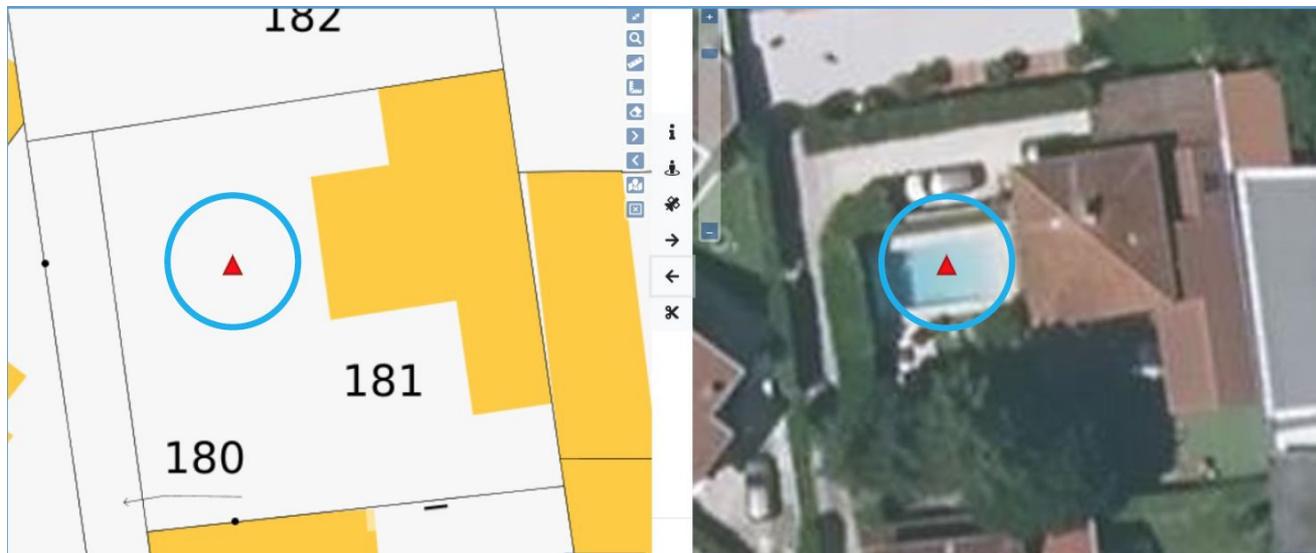


Cas d'étude — Projet Foncier Innovant (DGFIP)

**Du sol aux toits.
Quand une IA d'État tient compte
du territoire.**

Foncier
Innovant
DGFIP

2017 - ...



Cas d'application de géo-traitements concluants



Cas non concluants



... vers un
Report au plan
automatisé
DGFIP



PAR JUSTINE CARASSOU
14 octobre 2020, 12:24

L'IA vecteur de simplicité et d'efficacité du service public

Intelligence artificielle

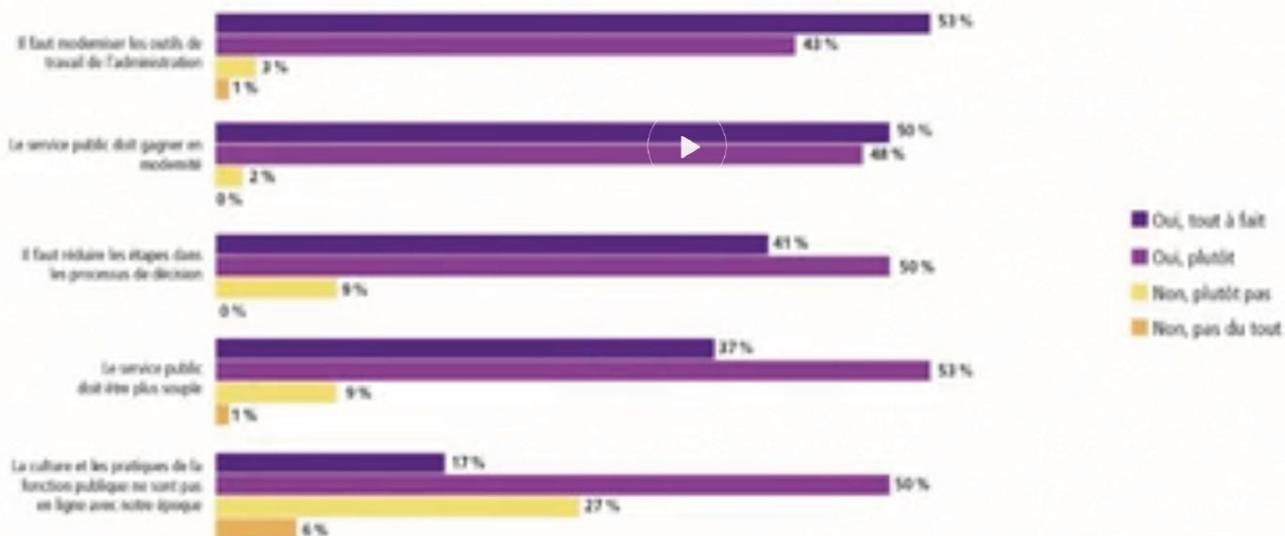
services publics

Yannick Girault, Directeur du service Cap Numérique, DGFIP et Gabriel Bellenger, Directeur Exécutif, Conseil en management pour les activités Santé et Secteur Public en Europe d'Accenture nous expliquent comment avec l'IA, les services publics peuvent devenir plus simples et efficaces et mieux répondre aux attentes partagées des citoyens et des agents.

accenture

La volonté de voir la fonction publique gagner en modernité et en agilité

L'attitude vis-à-vis de la fonction publique



PAR JUSTINE CARASSOU
14 octobre 2020, 12:24

L'
se

Inte

Yann
mani
les s
city

(2017-...) NOUVELLE EPREUVE

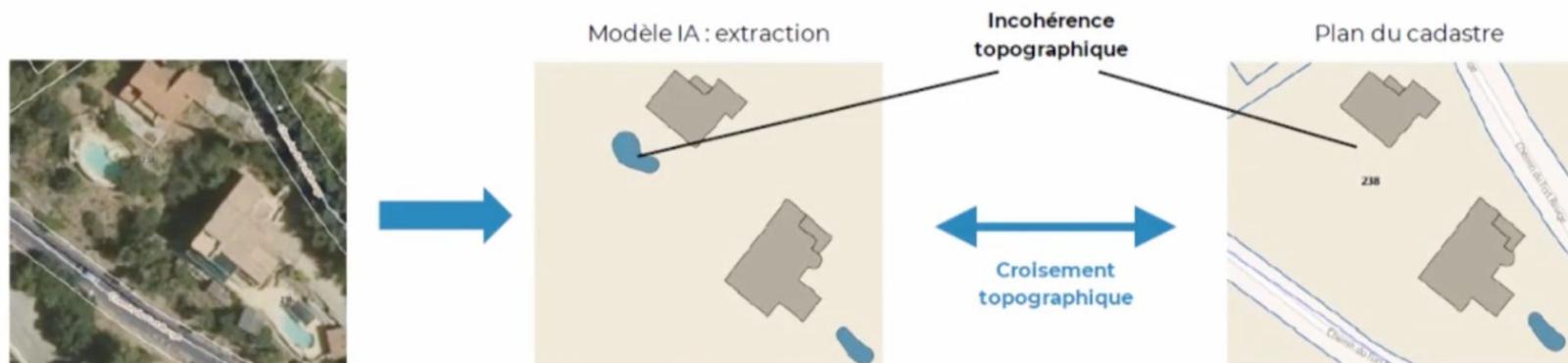
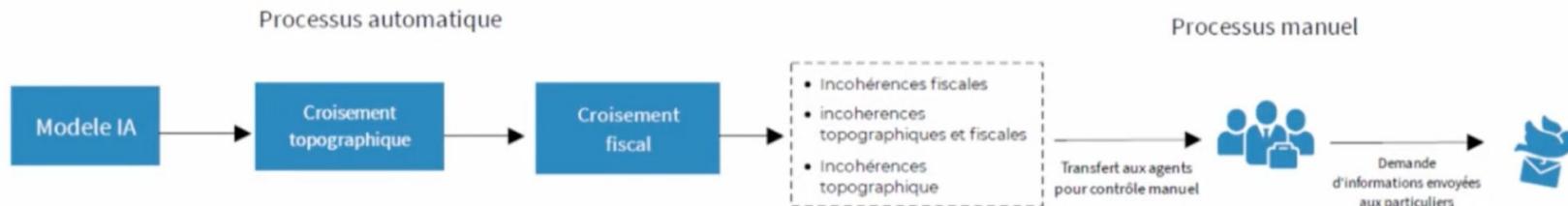
→ **Automatisation de la détection d'objets** (piscines, bâtiments isolés, extensions) & **mise à jour du plan cadastral avec des techniques d'intelligence artificielle**

“On est un peu sur **une révolution et sur une transformation profonde de nos missions et de nos métiers**. Et tout ça à un instant T, on fait tout en même temps.

En fait, **la dématérialisation, le recours à l'IA—ou à une technologie innovante au sens large—pour identifier les biens qui doivent faire l'objet d'une nouvelle évaluation, les biens qui doivent faire l'objet d'une représentation graphique sur le plan...** tout ça arrive au même moment. Et donc socialement, sociologiquement, ça a des impacts assez massifs sur l'organisation du travail, sur les missions au quotidien. Et tout ça arrive, entre guillemets un peu brutalement.”

(F., directeur du bureau du cadastre)

Processus du projet



Polygonal Building Extraction by Frame Field Learning

Nicolas Girard¹ Dmitry Smirnov² Justin Solomon² Yuliya Tarabalka³

¹Université Côte d'Azur, Inria ²Massachusetts Institute of Technology ³LuxCarta Technology

Abstract

While state of the art image segmentation models typically output segmentations in raster format, applications in geographic information systems often require vector polygons. To help bridge the gap between deep network output and the format used in downstream tasks, we add a frame field output to a deep segmentation model for extracting buildings from remote sensing images. We train a deep neural network that aligns a predicted frame field to ground truth contours. This additional objective improves segmentation quality by leveraging multi-task learning and provides structural information that later facilitates polygonization; we also introduce a polygonization algorithm that utilizes the frame field along with the raster segmentation. Our code is available at <https://github.com/Lydzorn/Polygonization-by-Frame-Field-Learning>.

1. Introduction

Due to their success in processing large collections of noisy images, deep convolutional neural networks (CNNs) have achieved state-of-the-art in remote sensing segmentation. Geographic information systems like Open Street Map (OSM) [30], however, require segmentation data in vector format (e.g., polygons and curves) rather than raster format, which is generated by segmentation networks. Additionally, methods that extract objects from remote sensing images require especially high throughput to handle the volume of high-resolution aerial images captured daily over large territories of land. Thus, modifications to the conventional CNN pipeline are necessary.

Existing work on deep building segmentation generally falls into one of two general categories. The first vectorizes the probability map produced by a network



Figure 1: A frame field output by our network.

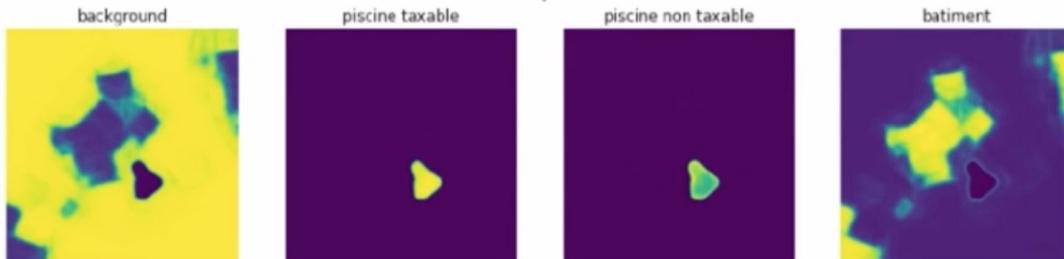
a posteriori, e.g., by using contour detection (marching squares [25]) followed by polygon simplification (Ramer-Douglas-Peucker [12, 13]). Such approaches suffer when the classification maps contain imperfections such as smoothed out corners, a common artifact of conventional deep segmentation methods. Moreover, as we show in Fig. 2, even perfect probability maps are challenging to polygonize due to shape information being lost from the discretization of the raster output. To improve the final polygons, these methods employ expensive and complex post-processing procedures. ASP polygonization [23] uses polygonal partition refinement to approximate shapes from the output probability map based on a tunable parameter controlling the trade-off between complexity and fidelity. In [45], a decoder and a discriminator regularize output probability maps adversarially. This requires computing large matrices of pairwise discontinuity costs between pixels and involves adversarial training, which is less stable than conventional supervised learning.

Another category of deep segmentation methods learns a vector representation directly. For example, CurveGCN [23] trains a graph convolutional network (GCN) to deform polygons iteratively and PolyMapper [21] uses a recurrent neural network (RNN) to predict vertices one at a time. While these approaches directly predict polygon parameters, GCNs and RNNs suffer from several disadvantages. Not only are they more difficult to train than CNNs, but also their output topology is restricted to simple polygons without holes—a serious limitation in segmenting complex buildings. Additionally, adjoining buildings with common walls are common, especially in city centers. CurveGCN and PolyMapper are unable to reuse the same polyline in adjoining buildings, yielding overlaps and gaps.

We introduce a building extraction algorithm that avoids the challenges above by adding a frame field output to a fully-convolutional network (see Fig. 1). While this has imperceptible effect on training or inference time, the frame field not only increases segmentation performance, e.g., yielding sharper corners, but also provides useful information for vectorization. Additional losses learn a valid frame field that is consistent with the segmentation. These losses regularize the segmentation, similar to [39], which includes



Deeplabv3+



REPUBLICQUE FRANÇAISE

Foncier Innovant

FRANÇOIS BILLOUES

Le report au plan

Opération manuelle de vectorisation



Opération automatique

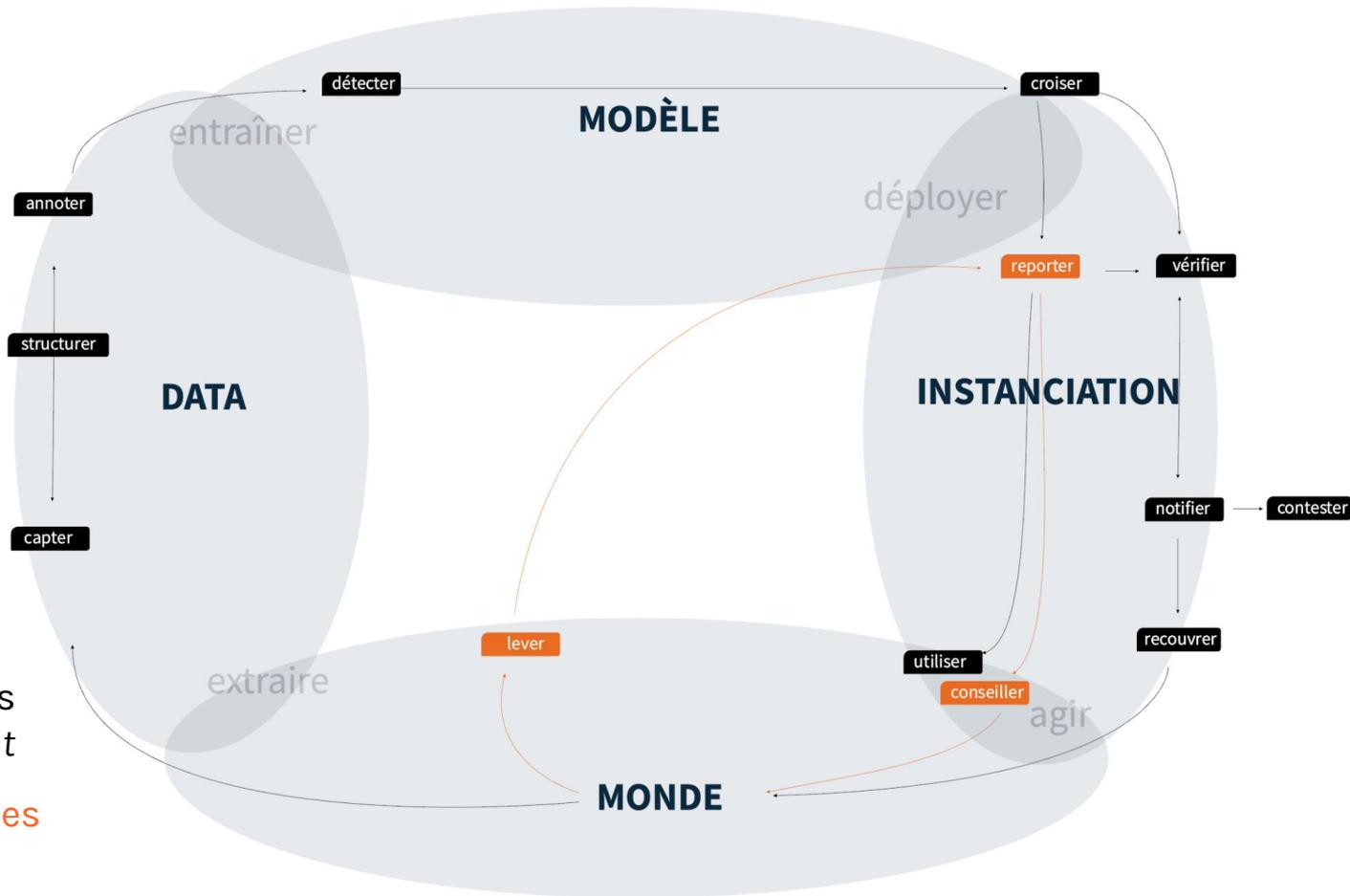


Présentation du projet Foncier Innovant
Yannis Tannier

30:28 / 04:27:18

The image shows a presentation slide comparing manual and automatic vectorization of an aerial photograph. The left side, labeled 'Opération manuelle de vectorisation', shows a red outline of a building complex with a jagged, irregular shape, indicating manual drawing. Below it is an icon of a group of people. The right side, labeled 'Opération automatique', shows the same red outline but with a much smoother and more precise boundary, indicating the use of automated software. Below it is a gear icon. The slide includes logos for the French Republic and 'Foncier Innovant', and a small video inset in the top right showing a man speaking. A footer contains the presenter's name 'Yannis Tannier' and a video player interface with a progress bar and controls.

Slide issue d'une présentation officielle du projet Foncier Innovant. Séminaire IA, « Imagerie aérienne et report au plan » organisé par la délégation à la transformation numérique de la direction générale des Finances publiques (DTNum), 29 juin 2022.



La boucle de l'IA

→ les tâches induites par le Foncier Innovant

→ les tâches actuelles des géomètres



“On voit bien qu’il y a tout un écosystème qui se développe dont le but est de changer des relations existantes par d’autres.”

(H., journaliste, participant au workshop - nov. 2022)

établir standards: quel-acquis? Automatiser

Degré d'opacité / interaction avec le modèle

PROBLÉMATISATION

Vérification / Travail / Humain / DGFIP

étape pour représenter le monde (Clobbar)

Rapport de puissance

l'usage humain

Standards innovatifs

Quel est ce qu'on veut employer? quel est ce qu'on veut accomplir? le quel? Suivi-jc responsable

QUALITE

Biais

Employer / Habiter / Collecte de données

INSTANTIATION

MAÎTRISE des décisions HUMAINES

CLASSIFICATION

maires / permis de construire / citoyens / notaires? / territoires / parcelles / bâtiments / propriétaires / revenus

Bäcker and Stan

l'usage humain / contribuable / quel est ce qu'on veut accomplir? le quel? Suivi-jc responsable

NOT FRONT / work / kees

opague

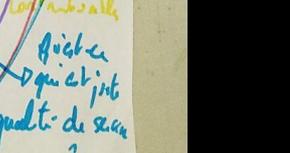
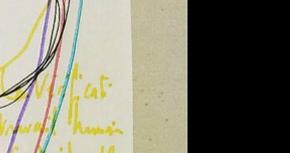
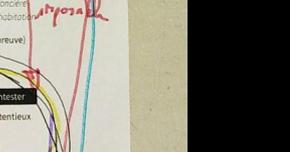
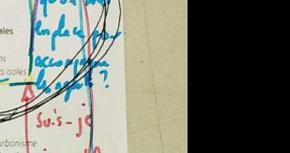
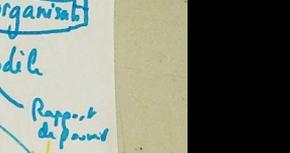
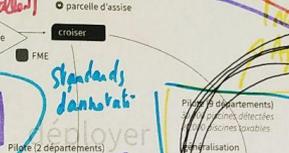
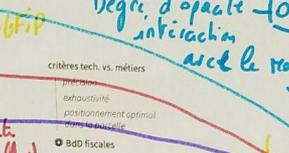
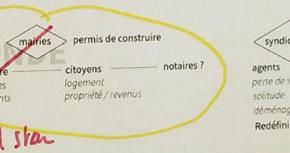
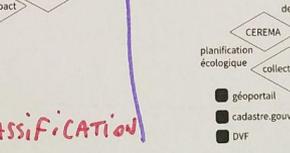
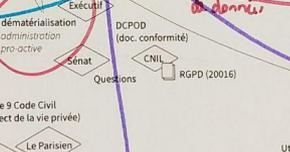
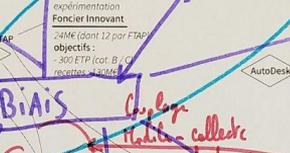
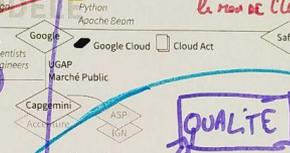
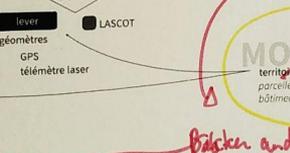
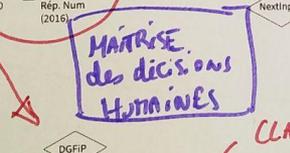
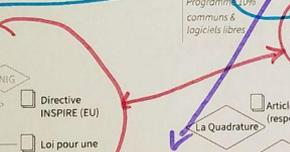
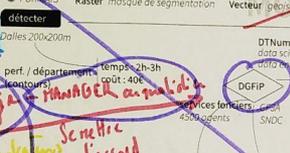
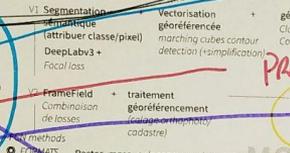
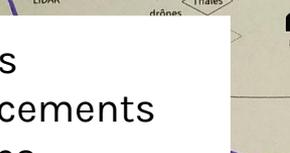
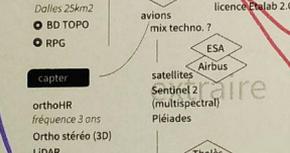
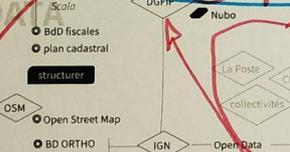
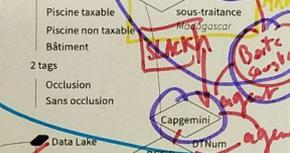
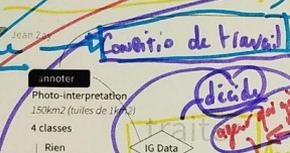
complexité

DATA

faire

MOISE

1990-2022



→ les soucis des nouveaux agencements socio-techniques

	tenir compte	soucis associés
Qu'est ce qui compte ?	→ cohérence des bases de données	justice fiscale, économie
	→ découplage représentation graphique et connaissances bâtiments	changement des critères (précision) et transfo. expertise métier
	→ mises à jour régulières	données, traduction décisions politiques en choix techniques
	→ open data, réemploi données	économie politique, sécurité
Comment on tient compte ?	→ déplacement, division et invisibilisation du travail	4 pôles de travail, mutation des tâches économie politique, relation citoyen / usager plateforme
	→ élaborer une <i>vérité terrain</i>	travail (annotation: déplacement géographique, mauvaise qualité), perte de technicité / connaissance terrain, traduction technique des catégories
	→ une chaîne plus longue et plus distribuée	absence d'évaluation (coûts, efficience), vulnérabilités, privatisation
	→ compétition institutionnelle	silos (duplication des efforts), manque d'articulation avec les autres pol.publ. sur le territoire (ZAN, etc.)
Pour quoi ?	→ le monde laboratoire	économie politique (coûts, marchandisation données...), inversion du rapport fins/moyens, absence de pensées des effets (pensée par les risques), maintenance / durabilité ?
	→ état plateforme, dématérialisation	perte de relations au service public, aux citoyens et au terrain, contrôle et inversion de la charge de la preuve
	→ compenser des politiques publiques impopulaires	tensions entre échelles territoriales, consentement à l'impôt



Aluminum

Al

China

6



Manganese

Mn

Germany

10



Antimony

Sb

China

14



Zinc and Arsenic

Zn+As

China

18



Cadmium

Cd

China

22



Coal

C

China

26



Gold, Silver, and Titanium

Au+Ag+Ti

China

30



Beryllium

Be

China

7



Tungsten

W

China

11



Cobalt

Co

China

15



Lithium

Li

China

19



Lithium

Li

China

23



Copper

Cu

China

27



Copper

Cu

China

31



Gallium

Ga

China

8



Neodymium

Nd

China

12



Oil shale

China

16



Magnesium

Mg

China

20



Sulfur

S

China

24



Iron

Fe

China

28



Phosphorus

P

China

32



Boron

B

China

9



Potassium

K

China

13



Silicon

Si

China

17



Chromium

Cr

China

21



Tin and Tantalum

Sn+Ta

China

25



Rare Earth Elements

REE

China

29



Mercury

Hg

China

33



Technology

OpenAI CEO Altman says at Davos future AI depends on energy breakthrough

By Reuters

January 16, 2024 6:39 PM GMT+1 · Updated 3 months ago



Sam Altman, CEO of OpenAI, attends the Asia-Pacific Economic Cooperation (APEC) CEO Summit in San Francisco, California, U.S. November 16, 2023. REUTERS/Carlos Barria/File Photo [Purchase Licensing Rights](#)

DAVOS, Switzerland, Jan 16 (Reuters) - OpenAI's CEO Sam Altman on Tuesday said an energy breakthrough is necessary for future artificial intelligence, which will consume vastly more power than people have expected.

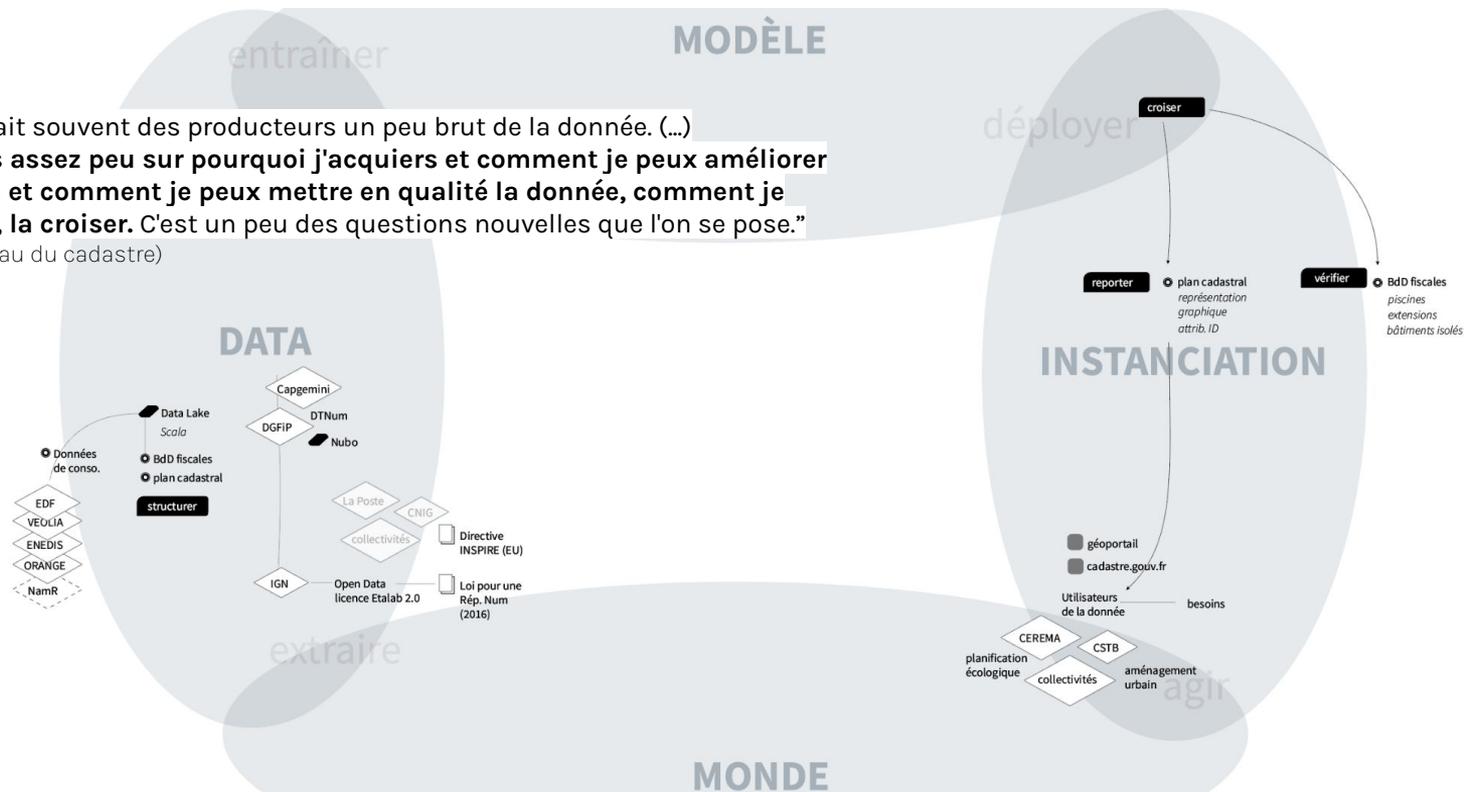
Speaking at a Bloomberg event on the sidelines of the World Economic Forum's annual meeting in Davos, Altman said the silver lining is that more climate-friendly sources of energy, particularly nuclear fusion or cheaper solar power and storage, are the way forward for AI.

Qu'est-ce qui compte ?

→ Les “couloirs de valorisation” de la donnée

“À la DGFiP, on était souvent des producteurs un peu brut de la donnée. (...) Je m'interrogeais assez peu sur pourquoi j'acquies et comment je peux améliorer cette acquisition et comment je peux mettre en qualité la donnée, comment je peux la valoriser, la croiser. C'est un peu des questions nouvelles que l'on se pose.”

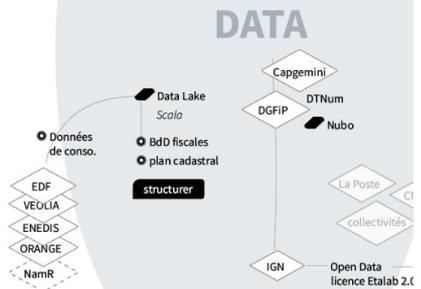
(F., directeur du bureau du cadastre)



Qu'est-ce qui compte ?

→ Les “couloirs de valorisation” de la donnée

“À la DGFiP, on était souvent des producteurs un peu. Je m'interrogeais assez peu sur pourquoi j'acquiescerais cette acquisition et comment je peux mettre en œuvre comment je peux la valoriser, la croiser. C'est un peu des questions” (F., directeur du bureau du cadastre)



DATA

extraire

MODÈLE



Sébastien Soriano

@sorianotech

...

À dispo pour aider @dgfip_officiel 🙌 Quantité de startups talentueuses savent exploiter les vues aériennes de @IGNFrance 🇫🇷, qui sont en open data, et sans recours aux GAFAs. On peut vous mettre en relation avec celles que l'on a déjà identifiées via notre programme IGNfab.



@canardenchaine @canardenchaine · 10 août

DEMAIN DANS LE CANARD

Google au service du fisc pour taxer les piscines et autres constructions non déclarées

12:26 AM · 12 août 2021 · Twitter for iPhone

› fiscales
› inés
› missions
› iments isolés

Comment on tient compte ?

→ **Élaborer une vérité terrain** (*ground truth*)

“Et puis pour arriver à dématérialiser, il y a un certain nombre de fondations à poser telles que **la mise en place de référentiels**, notamment le référentiel des locaux.”

(F., directeur du bureau du cadastre)

D’abord les habitations devront être classées selon quatre sous-groupes (maisons individuelles, appartements situés dans les immeubles collectifs, dépendances isolées et locaux d’habitation qui présentent des caractéristiques exceptionnelles). Un décret en Conseil d’État viendra ensuite décomposer les trois premiers sous-groupes en plusieurs catégories de locaux selon leur consistance (pour les maisons individuelles et les appartements) ou leur utilisation (pour les dépendances isolées).

sources :

< Rapport de la **Cour des Comptes** sur la révision des valeurs locatives, 2023

√ Slides de présentation de foncier innovant, séminaire DGFiP, Juin 2022

Pilote Piscine et Bati : Labellisation

- 150 zones de 1km² sélectionnées
- 4 classes :
 - 0 : rien
 - 1 : piscine taxable
 - 2 : piscine non taxable
 - 3 : bâtiment
- Tag avec occlusion ou sans occlusion
- 2 formats : Vecteur et Raster

The slide shows a workflow from a satellite image of a residential area to a vector map where buildings are highlighted in black, representing the 'bâtiment' class.

Le report au plan

Démonstration de la preuve de faisabilité, mais les contraintes sont fortes

- Un contour précis : des critères métiers stricts
- Un positionnement optimal de la forme au sein de la parcelle d'assise : des traitements de géoréférencement des contours prévus

	Critère n°1	Critère n°2	Critère n°3	Critère n°4	Critère n°5 et n°6	Critère n°7
Critère	Comparaison du nombre de sommets	Ecart de distance entre sommets homologues	Ecart surfacique	Taux de recouvrement	N°5 : Ecart angulaire entre sommets homologues N°6 : Ecart angulaire entre segments homologues	Comparaison des périmètres
Seuil	Aucun écart	50cm	10%	93%	L'angle ABC est formé par les segments [BA] et [BC] Si [BA] et [BC] > 1m Si 1m < [BA] et [BC] < 2m Si [BA] < 1m et [BC] > 2m	5%

The slide includes a video thumbnail of a speaker and a map showing the application of these criteria to a specific parcel.

Comment on tient compte ?

→ **Élaborer une vérité terrain** (*ground truth*)

Menu

Le Parisien

🔍 👤

S'abonner

Taxe foncière sur les piscines : comment Daniel a fait plier le fisc pour ne pas la payer

Le Trésor public réclamait à un habitant de l'Oise de déclarer sa piscine qui avait été repérée par l'exploitation d'images satellites. Après des semaines de bataille et un article que nous avons consacré à cette histoire, le service foncier des impôts des particuliers de Beauvais a tranché : il ne sera pas imposé.



Daniel avait notamment mis en avant auprès du fisc le fait que sa piscine ne reste pas en place plus de trois mois par an. DR

Par [Maxime Gayraud](#)

Le 24 juillet 2023 à 16h35, modifié le 27 juillet 2023 à 08h14

Son histoire, que nous avons [racontée la semaine dernière](#), a été une des plus partagées et des plus commentées de ce mois de juillet par les lecteurs du Parisien - Aujourd'hui en France. Et elle vient de connaître un épilogue heureux pour le contribuable concerné. Daniel (*le prénom a été changé*), un habitant de l'Oise, ne verra pas le montant de sa taxe foncière grimper à cause de la piscine installée dans son jardin.

Pourtant, depuis le mois dernier, le fisc lui réclamait de déclarer son bassin, ce qui aurait fait grimper la valeur locative de sa maison et alourdi son imposition. Sa piscine avait en effet fait l'objet d'une détection automatique dans le cadre du programme baptisé « Foncier innovant ». Grâce à un logiciel

Comment on tient compte ?

→ **Élaborer une vérité terrain** (ground truth)

Colonne	Titre	Contenu
1	N°	Numéro d'ordre
		Année de la mise à jour
2	MAJ	Il est précisé que l'année de mise à jour n'entraîne pas automatiquement l'application de l'exonération prévue à l'article 1383 du code général des impôts (CGI).
3	N° de PEV	Numéro de partie d'évaluation (le cas échéant)
4	Type de local	Nature du local selon les valeurs ci-dessous : - 1 : maison ; - 2 : appartement ; - 3 : local n'ayant que des dépendances
5	Commune absorbée	Code de la commune absorbée (le cas échéant)
6	Sect.	Identification parcellaire du lieu de situation du local
7	N° plan	Remarque : le numéro de bâtiment est un numéro interne à l'administration
8	Bât.	
9	Invariant	Le numéro invariant est composé de 7 caractères numériques. Il s'agit d'un numéro d'identification interne à l'administration
10	Adresse Local	Adresse de situation du local
11	Propriétaire	Nom ou dénomination du propriétaire du local Dans le cas où plusieurs personnes sont propriétaires, une seule personne est restituée : il s'agit du destinataire de l'avis d'imposition.
12	Nature	Description synthétique de la nature du local
13	Ancienne VL	Valeur locative 70 du local avant mise à jour ou changement d'affectation
14	Nouvelle VL	Valeur locative 70 du local après mise à jour ou changement d'affectation
15	Affectation	Affectation de chaque partie d'évaluation du local à l'aide du code HAB : habitation
		Pour un local d'habitation : Nombre de pièces de la partie d'évaluation du local (cuisines, salles d'eau et annexes non comptées)
		Pour les dépendances : Nature de la dépendance que constitue la partie d'évaluation. Les codes suivants sont utilisés : - GA : garage ; - GP : garage-parking ; - PK : parking ; - BX : box ; - CL : cellier ; - CV : cave ; - RS : remise ; - BC : bûcher ; - BD : buanderie ; - GR : grenier ; - DC : dépendances locaux communs ; - PI : pièce indépendante ; - PA : dépendance pur agrément ; - CD : chambre domestique ; - GC : grenier-cave ;
16	Nbre de Pièces	

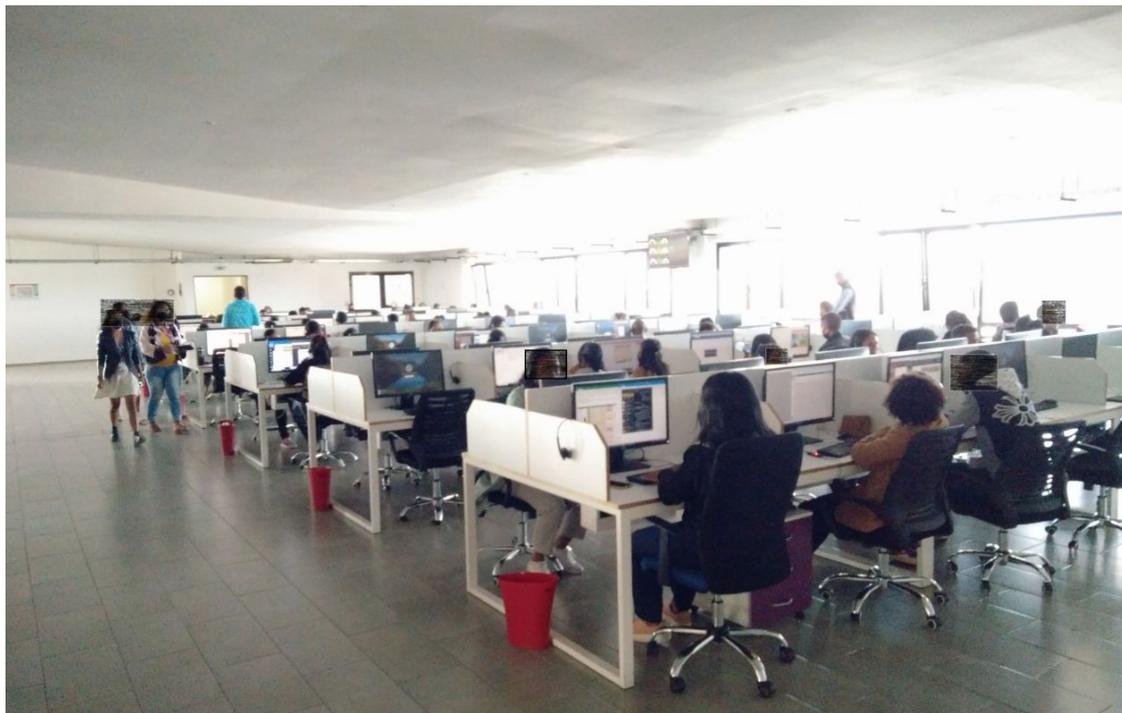
Colonne	Titre	Contenu
17	Surface réelle	Surface réelle, mesurée au sol, de la partie principale ou de la dépendance.
		Catégorie prise en compte pour la partie d'évaluation (article 324 H de l'annexe III au CGI). Elle peut prendre les valeurs suivantes :
		Pour la partie principale à usage d'habitation ou professionnel (la lettre M symbolise les catégories intermédiaires, par exemple 4M entre les catégories 4 et 5) : - 1 : grand luxe ; - 2 : luxe ; - 3 : très confortable ; - 4 : confortable ; - 5 : assez confortable ; - 6 : ordinaire ; - 7 : médiocre ; - 8 : très médiocre.
18	Catégorie	Pour les dépendances ordinaires isolées (garages, remises,...) : - A : bonne ; - B : passable ; - C : médiocre ; - D : mauvaise. Pour les dépendances d'agrément (piscines) : - AA : bonne ; - BA : passable ; - CA : médiocre ; - DA : mauvaise.
19	Équivalences superficielles	Conversion en m ² des éléments de confort et d'équipement du local déclarés par le propriétaire (eau, chauffage central, électricité...). Cette conversion est établie à partir d'un barème d'équivalences superficielles pour chaque élément de confort (article 324 T de l'annexe III au CGI).
20	Correctif	Correctif d'ensemble, défini à partir de l'état d'entretien de la partie d'évaluation (article 324 Q de l'annexe III au CGI), de sa situation (article 324 R de l'annexe III au CGI) et, le cas échéant, d'un correctif d'ascenseur (article 324 S de l'annexe III au CGI) permettant d'ajuster la valeur locative.
21	Tarif	Donnée permettant de distinguer différentes zones d'une commune ayant chacune nécessité une classification particulière des locaux. Elle n'est pas éditée pour les communes à série tarif unique.
22	Motif mise à jour	Donnée, facultative, permettant de préciser les éléments ayant conduit à la nouvelle évaluation.

“La question est : **est-ce que visuellement l'information est là ?** Est-ce que j'arrive à l'œil nu à placer les frontières ? C'est à dire que si c'est faisable, si à l'œil nu, un humain arrive à faire la tâche, alors on peut imaginer qu'un réseau arrivera à le faire.”

(G., chercheur en IA)

Comment on tient compte ?

→ **Élaborer une vérité terrain** (*ground truth*)



^ Entreprise sous-traitant l'annotation de données, Tana, Madagascar - Crédits : Clément Le Ludec, Maxime Cornet, 2022

L'annotation

“Donc nous, notre **premier défi déjà, c'était de construire la base de connaissances**. Donc, il nous faut la typologie la plus large possible de types de piscine taxables, non-taxables. Pareil sur les bâtiments.

Donc recours à un prestataire externe avec regard de la DGFIP, mais là pareil un petit peu bancal au début parce qu'on n'avait pas toute la connaissance, tout le recul, **ça va très très vite. Il n'y a pas le temps, pas forcément de réactivité. On a une base de connaissances qui est faite en à peine deux mois**, qui permet de lancer l'apprentissage de l'algorithme.”

(F., directeur du bureau du cadastre)

“Et derrière cette d'opacité, eh bien il y a aussi du coup des **rapports de pouvoir**. Pour les annotateurs, on leur dit pas forcément à quoi leur travail sert et ça permet aussi de maintenir des conditions de travail plus basses.”

(C., sociologue)

Qu'est-ce qui compte ?

→ **La mise à jour et la fraîcheur des données**

Recommandation n° 6. (DGFIP) Poursuivre la dématérialisation des déclarations foncières et la numérisation des outils à la disposition de la DGFIP pour actualiser plus rapidement les bases d'imposition.

source :

< Rapport de la Cour des Comptes
sur la révision des valeurs locatives, 2023

Comment établir qu'il y a eu changement *pertinent* ? (topographique ou fiscal)

→ **difficultés de traduction** (quoi détecter ?)

→ **acquisition et traitement des données (images)**

“Et puis ça pose la question de la régularité, de la mise à jour du plan : tous les combien ça va être mis à jour ? **Les communes s'attendent à un peu plus que tous les quatre ans.** Puis ils sont en train de nous dire finalement ou nous enverra quand même sur le terrain. Mais on ne va pas attendre quatre ans pour voir ce qui manque pour aller sur le terrain ou vérifier que c'est bon.”

(A., agent géomètre de la DGFIP)

→ **perte d'équité fiscale ?**

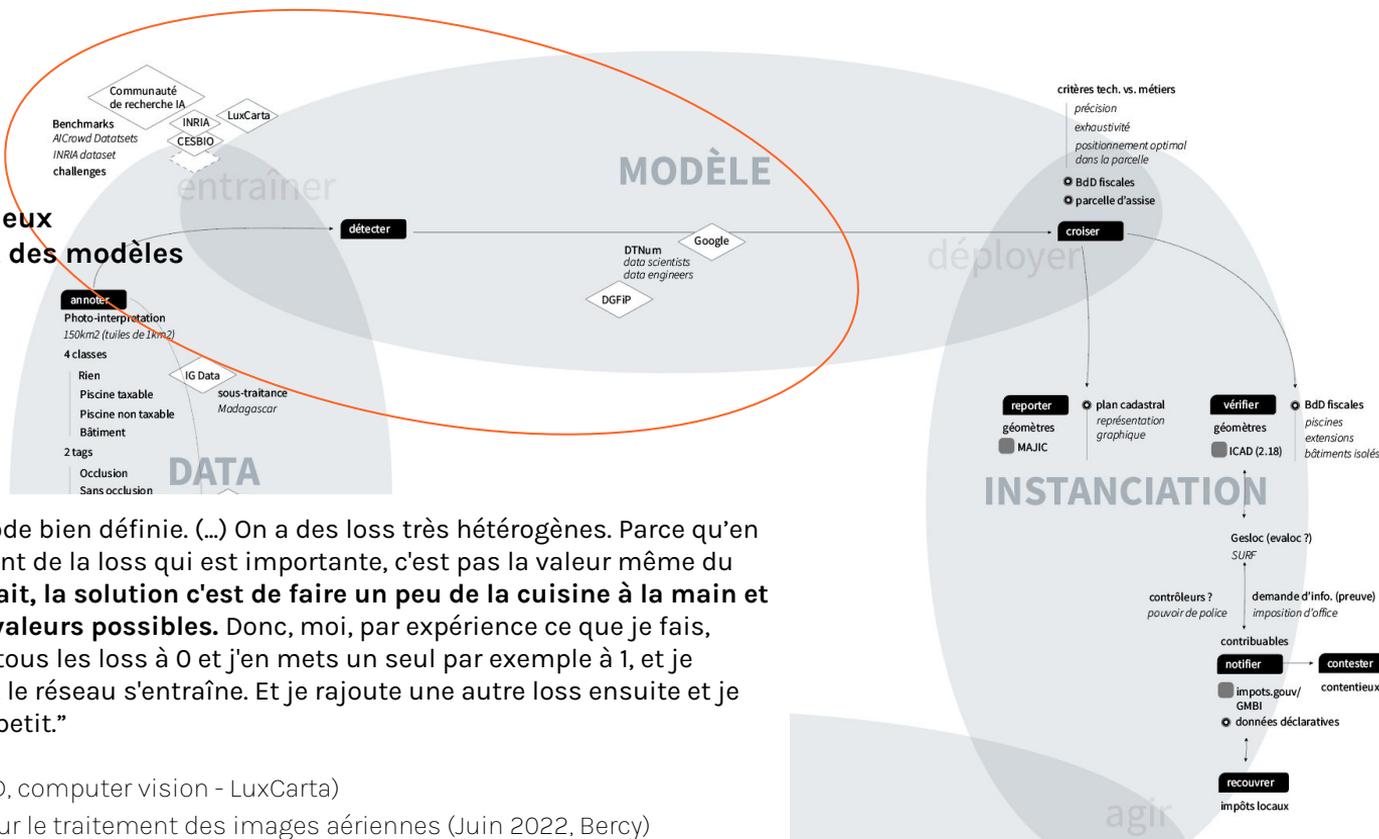
Temporalité importe : date du 1er janvier

Manque à gagner important (comparé à ce que rapporte la détection de piscines par commune par ex.)

Comment on tient compte ?

→ **Le déplacement et l'invisibilisation du travail**

Le travail laborieux d'entraînement des modèles



“Y a pas de méthode bien définie. (...) On a des loss très hétérogènes. Parce qu'en fait c'est le gradient de la loss qui est importante, c'est pas la valeur même du loss. **Et donc en fait, la solution c'est de faire un peu de la cuisine à la main et tester plusieurs valeurs possibles.** Donc, moi, par expérience ce que je fais, c'est que je mets tous les loss à 0 et j'en mets un seul par exemple à 1, et je regarde comment le réseau s'entraîne. Et je rajoute une autre loss ensuite et je les ajoute petit à petit.”

Nicolas Girard (PhD, computer vision - LuxCarta)

Séminaire DGFiP sur le traitement des images aériennes (Juin 2022, Bercy)

Pour quoi ça compte ?

→ **Le monde laboratoire**

« Si on fait un peu l'autopsie de ce cas là, finalement, on en vient rapidement à se demander pourquoi il y a un modèle d'IA dans ce cas d'application. (...) Les données aériennes de l'IGN qui sont produites tous les quatre ans, elles sont faisables à la main en fait, assez rapidement et pour beaucoup moins que 20 millions. (...) **Est ce que ce n'est pas un truc qui s'auto-justifie ? Est ce que toute cette construction là ne tient pas plus au fait de justifier le modèle que réellement d'aller taxer des piscines ?** Parce que finalement la taxation des piscines, si elles n'étaient pas traitée avant, c'est que tout le monde s'en fiche un peu, soyons honnêtes. Est ce que ce n'est pas juste parce que c'est un cas d'application simple à mettre en avant ? **C'est un peu le serpent qui se mord la queue, parce que finalement, on a traité le cas le plus simple en mettant beaucoup d'argent dedans. »**

(M., sociologue)

Pour quoi ça compte
→ Le monde laborat

« Si on fait un peu l'autop
pourquoi il y a un modèle
qui sont produites tous l
rapidement et pour beau
s'auto-justifie ? Est ce q
le modèle que réellemer
piscines, si elles n'étaien
honnêtes. Est ce que ce n
avant ? **C'est un peu le s**
cas le plus simple en m

(M., sociologue)

Accueil

Particuliers

Entreprises

Les ministres

Le ministère

Presse

Accueil du portail » Numérique » [\[Vidéo\] Journée de la donnée](#)

A+ A- 📄

[Vidéo] Journée de la donnée

29/11/2019

Le vendredi 29 novembre, a lieu la journée de la donnée parrainée par Cédric O, secrétaire d'État auprès du ministre de l'Économie et des Finances et du ministre de l'Action et des Comptes publics, chargé du Numérique.

[\[Vidéo\] Voir la journée du 29 novembre en différé](#)



Shaping AI | Accueil

medialab.github.io/ShapingAI/

Accueil Avoir prise... La boucle de l'IA Les récits de l'IA Les soucis de l'IA Un système d'IA

SciencesPo
MEDIALAB

**PRISES , MÉ.PRISES ,
EM.PRISES , DÉ.PRISES ,
RE.PRISES...**

**Les formes de participation
au développement de l'IA
en France**

Fin 2020.
Le cycle médiatique s'emballa sur l'intelligence artificielle, ses promesses et ses problèmes, usant et abusant d'une rhétorique sensationnaliste. Des innovations médicales révolutionnaires aux prophéties de la fin de l'humanité, des engagements éthiques des GAFAM aux mauvais calculs de ParcouSup, il est difficile de savoir quoi penser de ce phénomène, si ce n'est qu'il semble sur le point d'investir de manière indifférenciée tous les champs de la vie en tout point du globe.

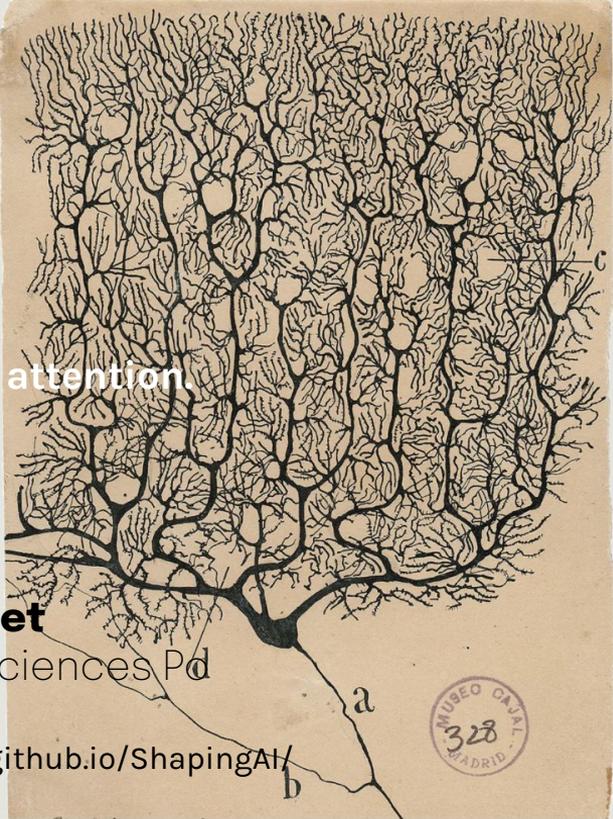
Difficile donc de naviguer sur cette mer de discours.

Dans le cas de l'intelligence artificielle, comme cela arrive de plus en plus à mesure que les systèmes se complexifient et se globalisent, une difficulté majeure tient à la résistance du phénomène à se laisser saisir. On ne sait

Merci de votre attention

Pauline Gourlet
médialab de Sciences Po

<https://medialab.github.io/ShapingAI/>



Dessin d'un neurone de Purkinje du cortex cérébelleux du chat par Santiago Ramón y Cajal (fin du 19e siècle). Source : Wikipedia

1/8