



LUND
UNIVERSITY



Inventors in 19th-Century France: Structuring and Linking Individual-Level Data through Machine Learning

YOUSSOUF MEROUANI, LUND UNIVERSITY



But really, it should be titled:
Here are some **new tools** for old problems and
strategies to avoid hallucinations.

The dataset I needed

- Covers the full period of interest: 1791-1899
- Unique IDs of each patent
- Unique IDs of individual inventors
- Descriptions of patents
- Technology Classification
- Would allow for construction of inventor-level data and patent-level data
- No missing data

The dataset I started with

- Covers the full period of interest: 1791-1899 ✗
- Unique IDs of each patent ✓
- Unique IDs of individual inventors ✗
- Descriptions of patents ✓
- Technology Classification ✗
- Would allow for construction of inventor-level data and patent-level data ✗
- No missing data ✗

The dataset I started with: INPI

ice Voir le dossier Ajouter à ma liste

Signaler une erreur

Cote du dossier	1BA185
Type de brevet	Brevet d'invention de 5 ans
Titre	parachute
Année de dépôt	1802
Déposant	GARNERIN Jeanne-Geneviève veuve, née LABROSSE
Observations	la déposante est l'épouse de André-Jacques GARNERIN, physicien aéronaute
Adresse du déposant	Paris (rue Plumet, faubourg Saint-Germain, Seine)/Paris (704, rue de Babylone, Seine)
Numéro de dépôt	Pas de numéro de dépôt entre 1791 et 1844
Date de dépôt	11/10/1802
Date de délivrance	16/11/1802
Observations générales	Le brevet porte le n°195 du Bureau des arts et manufactures. Le dossier contient une correspondance d'un litige sur le brevet publié dans la collection des Descriptions des machines et procédés.
Mot clé moderne	PARACHUTE
Mot clé historique	PARACHUTE
Classe	06. MARINE ET NAVIGATION/06.4 AEROSTATION, AVIATION

The dataset I started with: INPI

ice	Voir le dossier	Ajouter à ma liste	Signaler une erreur
Cote du dossier	1BA185		
Type de brevet	Brevet d'invention de 5 ans		
Titre	parachute		
Année de dépôt	1802		
Déposant	GARNERIN Jeanne-Geneviève veuve, née LABROSSE		
Observations	la déposante est l'épouse de André-Jacques GARNERIN, physicien aéronaute		
Adresse du déposant	Paris (rue Plumet, faubourg Saint-Germain, Seine)/Paris (704, rue de Babylone, Seine)		
Numéro de dépôt	Pas de numéro de dépôt entre 1791 et 1844		
Date de dépôt	11/10/1802		
Date de délivrance	16/11/1802		
Observations générales	Le brevet porte le n°195 du Bureau des arts et manufactures. Le dossier contient une correspondance d'un litige sur le brevet publié dans la collection des Descriptions des machines et procédés.		
Mot clé moderne	PARACHUTE		
Mot clé historique	PARACHUTE		
Classe	06. MARINE ET NAVIGATION/06.4 AEROSTATION, AVIATION		

The problems I have to solve:

- Extend the data to periods not covered
- Classify patent technology. Original data came with classifications for some of the patents.
- Figure out who's who: Linking individuals
- Individual-level information: Gender, HISCO, ...
- Fill in missing data for all periods

411004. Brevet de quinze ans, 15 décembre 1909; Société dite *'The North Eastern Marine Engineering Company Limited et M. Hunter (S.)*, représentés par Boramé et Julien, à Paris, boulevard Saint-Martin, n° 8. — Perfectionnements dans la construction des chambres de combustion des chaudières marines et autres et des foyers similaires.

411005. Brevet de quinze ans, 15 décembre 1909; Serraire (F.-E.), rue Jacques-Serraire, n° 10, à Nice (Alpes-Maritimes). — Roue à bandage élastique.

411006. Brevet de quinze ans, 17 décembre 1909; Jessup (J.-C.), représenté par la société Matray frères et C^{ie}, à Paris, boulevard Henri IV, n° 31. — Ressorts pour la confection de sommiers, de matelas, de sièges et d'autres objets rembourrés.

411007. Brevet de quinze ans, 17 décembre 1909; Morren (P.), représenté par Armengaud jeune, à Paris, boulevard de Strasbourg, n° 23. — Perfectionnement dans la suspension des véhicules et notamment des automobiles.

411008. Brevet de quinze ans, 18 décembre 1909; Paris (N.), à Brottes, par Chaumont (Haute-Marne). — Chausures à élastique unique, bout rapporté oblique.

411009. Brevet de quinze ans, 18 décembre 1909; Lacroix (J.), représenté par Rabilloud et fils, cours Morand, n° 10, à Lyon (Rhône). — Dispositifs destinés à faire reposer le poids d'un véhicule sur les axes des roues dans un sens oblique par rapport à l'appui vertical et dans de très bonnes conditions pour la traction.

411010. Brevet de quinze ans, 18 décembre 1909; Berthaud (J.), représenté par Germain, rue de l'Hôtel-de-Ville, n° 31, à Lyon (Rhône). — Dispositif de montage des appareils jette-bouts ou filières attaché-lames employés en filature.

411011. Brevet de quinze ans, 18 décembre 1909; Firme Gebrüder Heintze, représentée par Caius Danzer, à Paris, rue Richelieu, n° 5. — Corps de résonance pour instruments de musique, gramophones, etc.

411012. Brevet de quinze ans, 18 décembre 1909; Skita (A.), représenté par de Carsalade, à Paris, boulevard Beaumarchais, n° 2. — Procédé de production des terpènes à partir des huiles de térébenthine.

411013. Brevet de quinze ans, 18 décembre 1909; Bateman (R.-W.), représenté par de Carsalade, à Paris, boulevard Beaumarchais, n° 2. — Étau à griffes.

411014. Brevet de quinze ans, 18 décembre 1909; Giulietti (E.), représenté par Klotz, à Paris, boulevard de la Madeleine, n° 17. — Appareil mécanique amovible pour les changements de décors dans les théâtres.

411015. Brevet de quinze ans, 20 décembre 1909; Wewerka (H.), représenté par Blétry, à Paris, boulevard de Strasbourg, n° 2. — Procédé et dispositif pour la fabrication de corps creux en béton armé ou non armé, pierre artificielle ou argile au moyen de la force centrifuge et d'une pression intérieure.

411016. Brevet de quinze ans, 21 décembre 1909; Kühl (H.), représenté par Bonnet-Thirion, à Paris, boulevard Beaumarchais, n° 95. — Procédé de fabrication de ciment au moyen de laitier de haut fourneau.

411017. Brevet de quinze ans, 23 décembre 1909; Mumford (H.-T.), représenté par Beuttcher fils, à Paris, boulevard Saint-Martin, n° 39. — Outil pour laceter les courroies.

411018. Brevet de quinze ans, 23 décembre 1909; Twinberrow (J.-D.), représenté par Bert, à Paris, boulevard Saint-Denis, n° 7. — Ressorts coniques pour tampons de choc et autres applications.

411019. Brevet de quinze ans, 23 décembre 1909; Falk (L.), représenté par Bonnet-Thirion, à Paris, boulevard Beaumarchais, n° 95. — Procédé de fabrication de cérose.

411020. Brevet de quinze ans, 23 décembre 1909; Feige (A.), représenté par Armengaud jeune, à Paris, boulevard de Strasbourg, n° 23. — Jeu de billes avec distributeur automatique.

411021. Brevet de quinze ans, 23 décembre 1909; Hintermann (H.), représenté par Armengaud jeune, à Paris, boulevard de Strasbourg, n° 23. — Fermeture de sûreté pour récipients à liquides explosibles.



411004. Brevet de quinze ans, 15 décembre 1909; Société dite *The North Eastern Marine Engineering Company Limited* et *M. Hunter (S.)*, représentés par *Boramé et Julien*, à Paris, boulevard Saint-Martin, n° 8. — Perfectionnements dans la construction des chambres de combustion des chaudières marines et autres et des foyers similaires.



411004. Brevet de quinze ans, 15 décembre 1909; Société dite *The North Eastern Marine Engineering Company Limited* et
•NUM •TYPE •DATE •INV

M. Hunter (S.), représentés par *Boramé et Julien*, à Paris, boulevard Saint-Martin, n° 8. - Perfectionnement dans la
•INV •REP •REP •REP-ADD •DESCR

construction des chambres de combustion des chaudières marines et autres et des foyers similaires.



LUND
UNIVERSITY

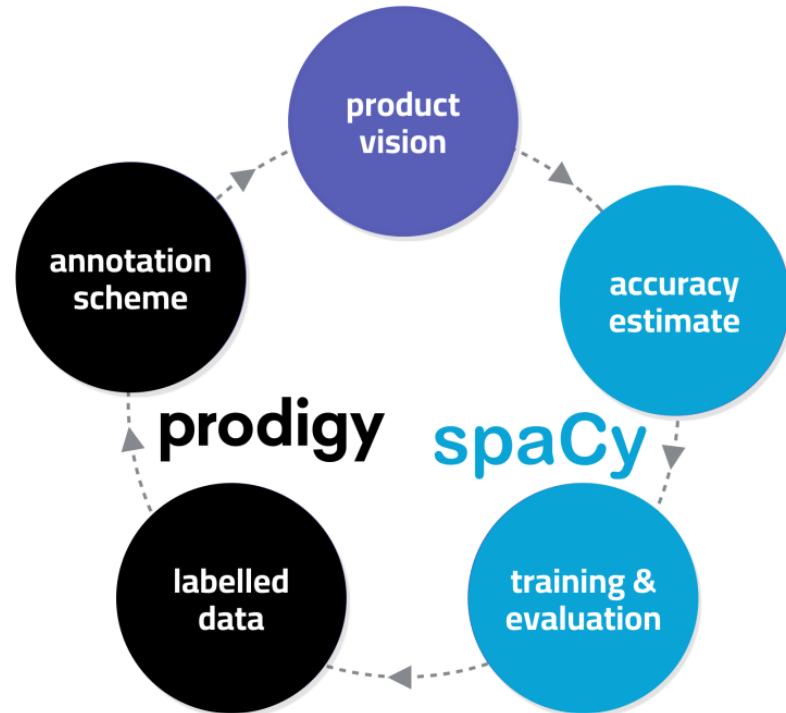
411004. Brevet de quinze ans, 15 décembre 1909; Société dite 'The North Eastern Marine Engineering Company Limited et M. Hunter (S.), représentés par Boramé et Julien, à Paris, boulevard Saint-Martin, n° 8. — Perfectionnements dans la construction des chambres de combustion des chaudières marines et autres et des foyers similaires.



411004. Brevet de quinze ans, 15 décembre 1909; Société dite The North Eastern Marine Engineering Company Limited et
•NUM •TYPE •DATE •INV

M. Hunter (S.), représentés par Boramé et Julien, à Paris, boulevard Saint-Martin, n° 8. - Perfectionnement dans la
•INV •REP •REP •REP-ADD •DESCR

construc- tion des chambres de combustion des chaudières marines et autres et des foyers similaires.



Problem 1: Assign Gender

140,624. Brevet de quinze ans, 14 janvier 1881; Novatzky de Angelis (M^{me}), représentée par Armengaud jeune, à Paris, boulevard de Strasbourg, n° 23. — Composition d'enduit destinée à rendre imperméables et imputrescibles le cuir, les étoffes, etc.

141,783. Brevet de quinze ans, 17 mars 1881; Cavallier (M^{me}), représentées par Desnos, à Paris, boulevard de Magenta, n° 11. — Mode de traitement des vidanges, boues de vidanges, liquides boueux, eaux vannes, matières grasses, etc.

140,714. Brevet de quinze ans, 21 janvier 1881; Andrieux (M^{me}), avenue de Neuilly, n° 31, à Neuilly (Seine). — Application de poudres métalliques ou autres sur les chapeaux de paille en général, et plus particulièrement sur ceux dits *chapeaux dentelle de paille*.



LUND
UNIVERSITY

Indistinguishable from noise after OCR:

M.

M'

M°

M''

MI

W..



LUND
UNIVERSITY

Problem 1: ~~Assign Gender~~. Fixing OCR-related problems

Classify



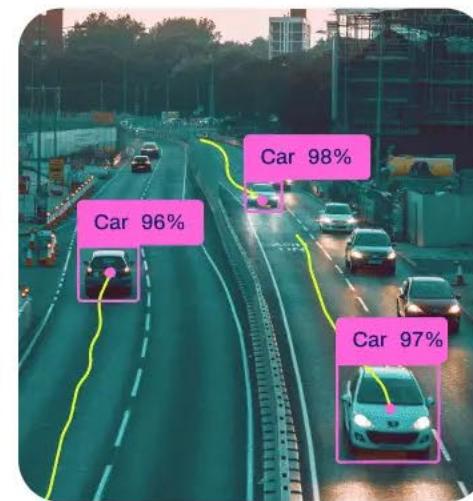
Detect



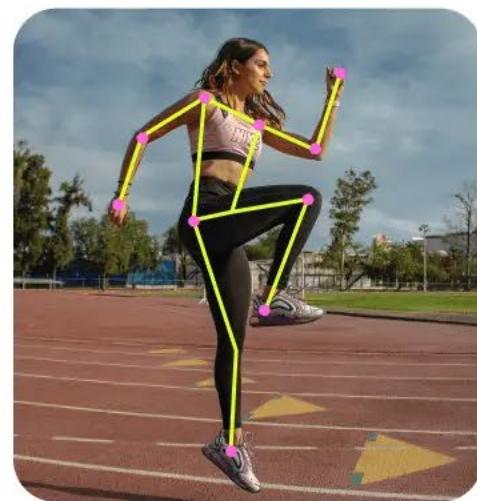
Segment



Track



Pose



LUND
UNIVERSITY



Export data and run three lines of code

```
from ultralytics import YOLO  
  
model = YOLO('yolov8n.pt')  
  
results = model.train(  
  
    data='data/annotated_patent_paragraphs.yaml',  
  
    name='yolov8n_marital_status'  
  
)
```

1896 P N°2112 N°2111-221-12-271003 mme 274925. Brevet de quinze ans, 2 février 1898; Rouyer (la dame), représentée par	1895 P N°1272 N°1714-1172-28-221125 231905. Brevet de quinze ans, 28 juin 1895; Content (Mme), place du boulevard du Mme 0.9	1902 N°2812/S. Brevet de quinze ans, 27 septembre 1898; Bathoni (la dame), rue des Trois-Couils, n° 16, à Bordeaux. — Le marqueur universel F. D.	189281. Brevet de quinze ans, 21 août 1898; Agier (Mme 0.8) rue Faithont, n° 30. — Robe dite sans couture.
1897 P N°2222 N°2122-12-25041 Mme 0.9 250415. Brevet de quinze ans, 21 septembre 1897; Verm Dorp (Mme veuve), repré	1895 P N°1272 N°1705-1812-21-25041 Mme 0.8 231850. Brevet de quinze ans, 1 ^{er} juin 1895; Mahalia (Mme), représentée par Chas-	18912. Brevet de quinze ans, 13 novembre 1896, Brun et Bourbaix, à Paris, avenue d'Orléans, n° 110. — Chariot dit : Panificateur pour l'armée.	187280. Brevet de quinze ans, 4 juillet 1892; M ^{me} Knight (M.-S.), représentée par Faber, à Paris, rue Joubert, n° 33. — Moteur à explosions.
1898 P N°2222 N°2111-222-12-270202 Mme 0.8 250284. Brevet de quinze ans, 4 novembre 1897; Lefèuvre (Mme), place des Halles,	1895 P N°2017 N°2054-1152-10-22 Mme 0.9 231903. Brevet de quinze ans, 29 juillet 1897; des Essards (Mme veuve), représentée par	18975/5. Brevet de quinze ans, 20 juillet 1897; Nelson (Mme L.), représentée par Blétry, à Paris, boulevard de Strasbourg, n° 2. — Corset.	1851211. Brevet de dix ans, 1 ^{er} mai 1895; Dubois (Mme), n° 2888 (E.-J.), à Paris, rue Saulnier, n° 13. — Plissé piqué indéplissable.
1898 P N°2112 N°2122-722-28-270422 mme 276489. Brevet de quinze ans, 1 ^{er} avril 1898; Vandamme (la dame veuve Henri),	1895 P N°7529. Brevet de quinze ans, 29 janvier 1894; M ^{me} Baumien (S.), représentée par Matray. — Protège-las.	183245. Brevet de quinze ans, 16 août 1891; M ^{me} Meyer (S.), représentée par Petiau, à Paris, avenue de la République, n° 44. — Ballons de « Pool-ball ».	18608. Brevet de quinze ans, 20 janvier 1892; Taggati (E.), représenté par M ^{me} Leymarie. — Mode d'attache de leviers de direction aux fusées.



LUND
UNIVERSITY

Mlle 0.9

140,714. Brevet de quinze ans, 21 janvier 1881; Andrieux (M^{me}), avenue de Neuilly, n° 31, à Neuilly (Seine). — Application de poudres métalliques ou autres sur les chapeaux de paille en général, et plus particulièrement sur ceux dits *chapeaux dentelle de paille*.



LUND
UNIVERSITY

Use it to construct frequency-type variables using rare-words, symbols, or figures directly from images.

Problem 2: Extract given names

Lesson learned: let's skip traditional OCR →
Multimodal LLMs to the rescue!



13

Brevet d'Invention

établi par la Loi du 7 Janvier 1791.

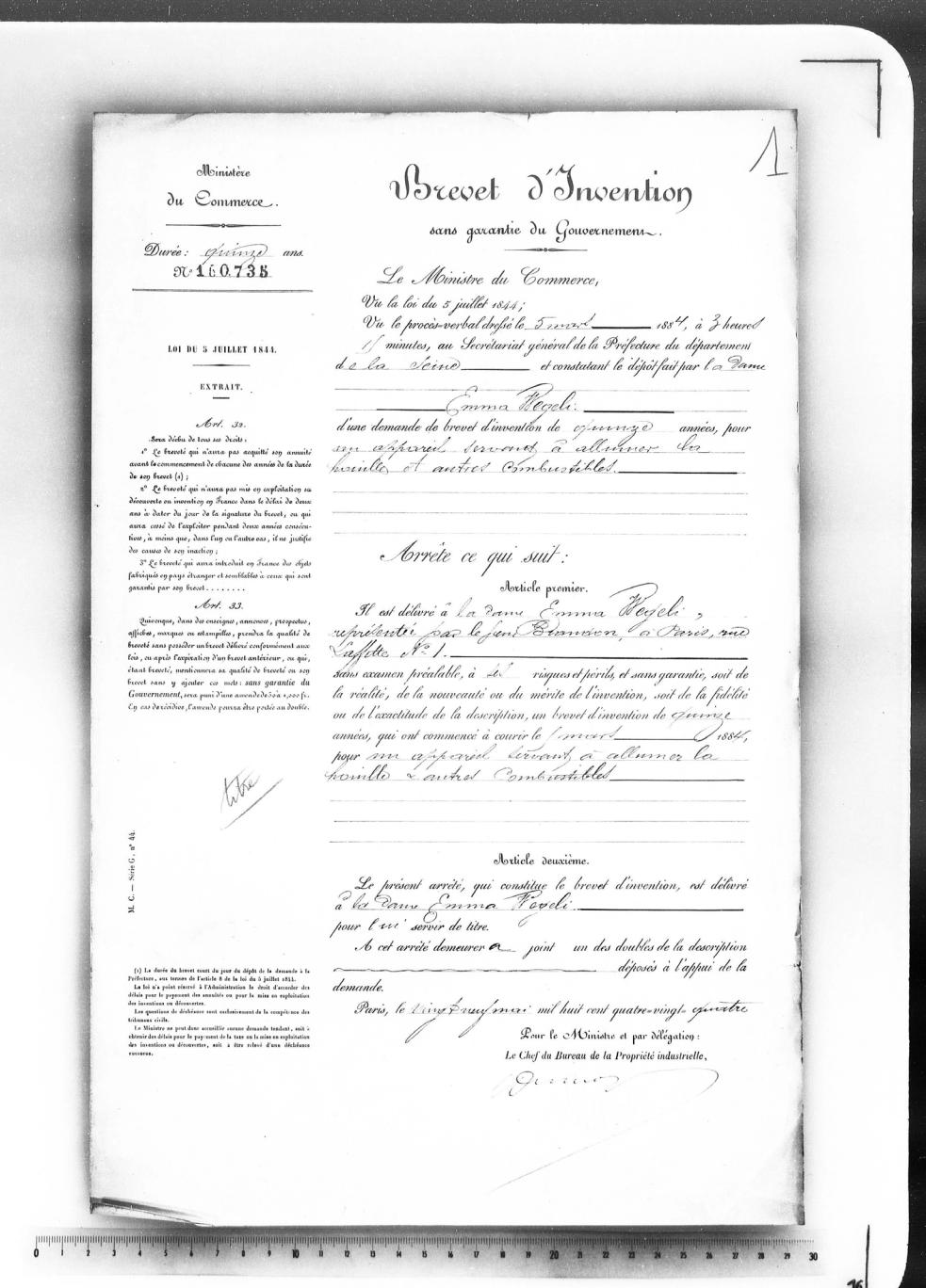
CERTIFICAT de demande d'un BREVET D'INVENTION,
délivré, en vertu de l'Arrêté des Consuls, du 5 Vendémiaire an 9,
au M. Jeanne-Geneviève Labrosse, femme Garnerin,
domicilié à Paris. département de la Seine,

Le Ministre de l'Intérieur

Vu la Petition présentée par Madame Jeanne Garnerin Labrosse, femme Garnerin, fondue de poruron, et demeurant à Paris rue Plumet, par laquelle elle expose que Son mari desire joindre des droits assurés par la Loi, aux auteurs d'inventions et découvertes en tout genre d'industrie, et en conséquence, obtenir, pour le terme de cinq années, un Brevet d'invention pour la machine nommée Parachute machine dont la 1^{re} Dame, au dit nous a déclaré que Son mari est l'auteur, ainsi qu'il résulte des pièces déposées au Secrétariat de la Préfecture du Département de la Seine, le Dix-neuf Vendémiaire de l'an ouze

Vu le mémoire descriptif de cette machine, ensemble une

135-462



LUND
UNIVERSITY



Brevet d'Invention

établi par la Loi du 7 Janvier 1791.

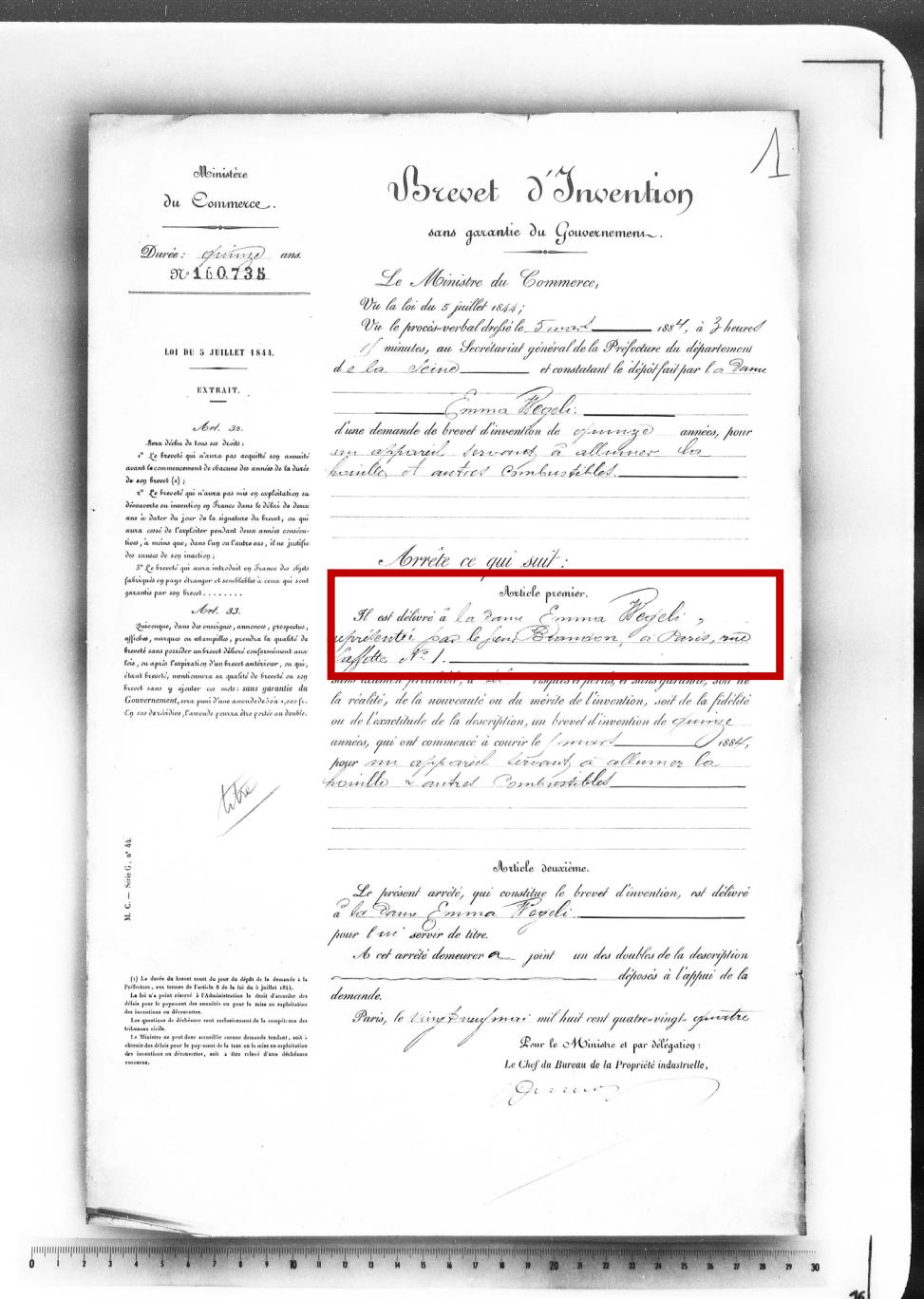
13

CERTIFICAT de demande d'un BREVET D'INVENTION,
délivré, en vertu de l'Arrêté des Consuls, du 5 Vendémiaire an 9,
au M. Jeanne-Geneviève Labrosse, femme Garnerin,
domicilié à Paris. département de la Seine.

Le Ministre de l'Intérieur

Vu la Petition présentée par Madame Jeanne Garnerin Labrosse, femme Garnerin, fondée de pouvoir, et devant à Paris rue Plumet, par laquelle elle expose que Son mari desire joindre des droits assurés par la Loi, aux auteurs d'inventions et découvertes en tout genre d'industrie, et en conséquence, obtenir, pour le terme de cinq années, un Brevet d'invention pour la machine nommée Parachute machine dont la 1^{re} Dame, au dit nous a déclaré que Son mari est l'auteur, ainsi qu'il résulte des pièces déposées au Secrétariat de la Préfecture du Département de la Seine, le Dix-neuf Vendémiaire de l'an ouze
Vit le mémoire descriptif de cette machine, ensemble une

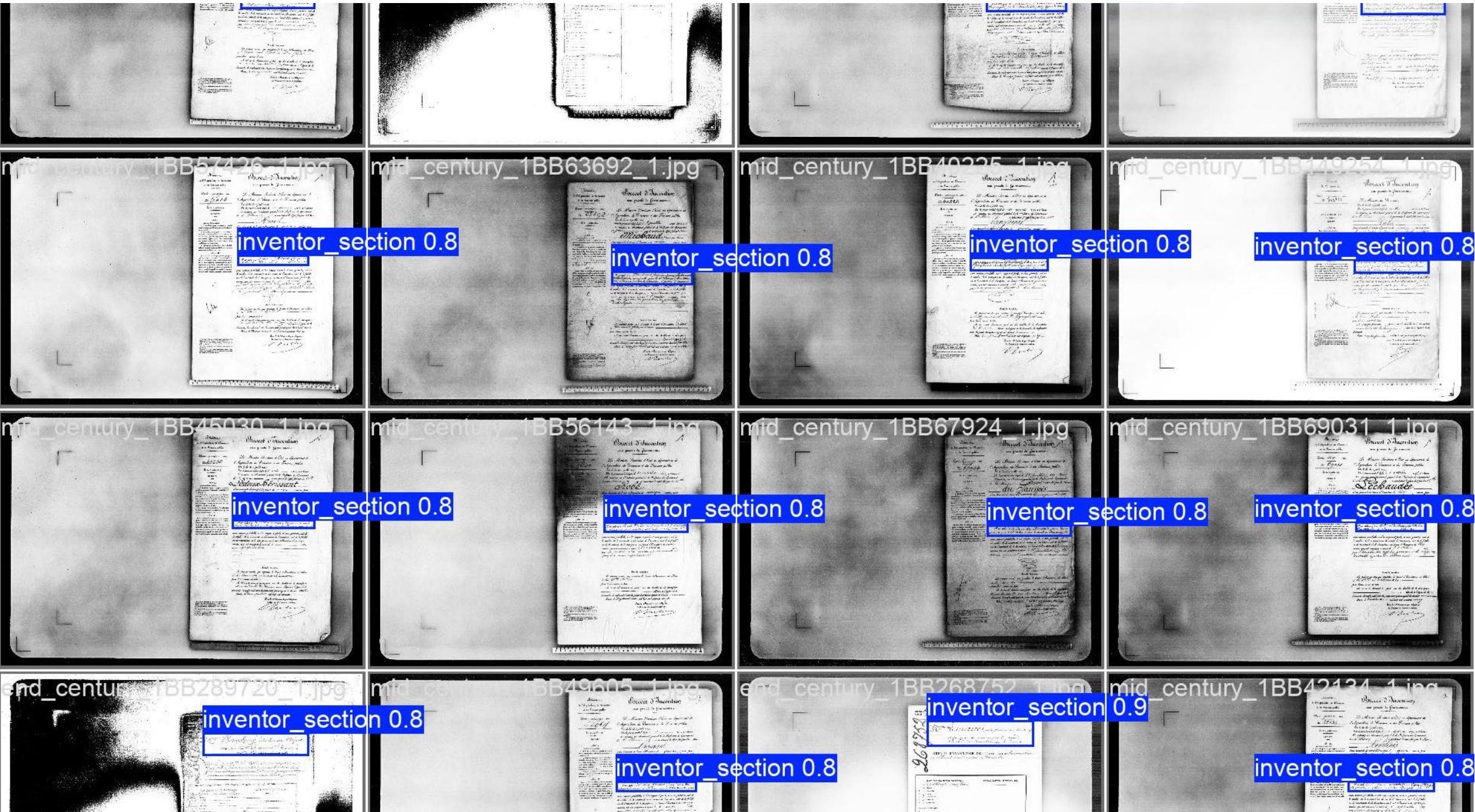
135-462



LUND
UNIVERSITY



Segment Predictions of the YOLO model



Closed models are easy to work with due to APIs. But so are Open models thanks to  LLM, which works great with Qwen-2.5-VL-7B

All vLLM needs for transcription:

- Hardware that is appropriate for your model of choice
 - 3B parameters fit inside an 8GB VRAM GPU.
 - No model for your language? Train it for free using Unsloth.
- A table with the location of your images (cloud or local).
- A prompt where you kindly ask the model to carefully transcribe and structure the text.

Avoid hallucinations and lower error-rate by forcing
two models to either agree or give up.

Large Language
Model

predict

profit?

Large Language
Model A

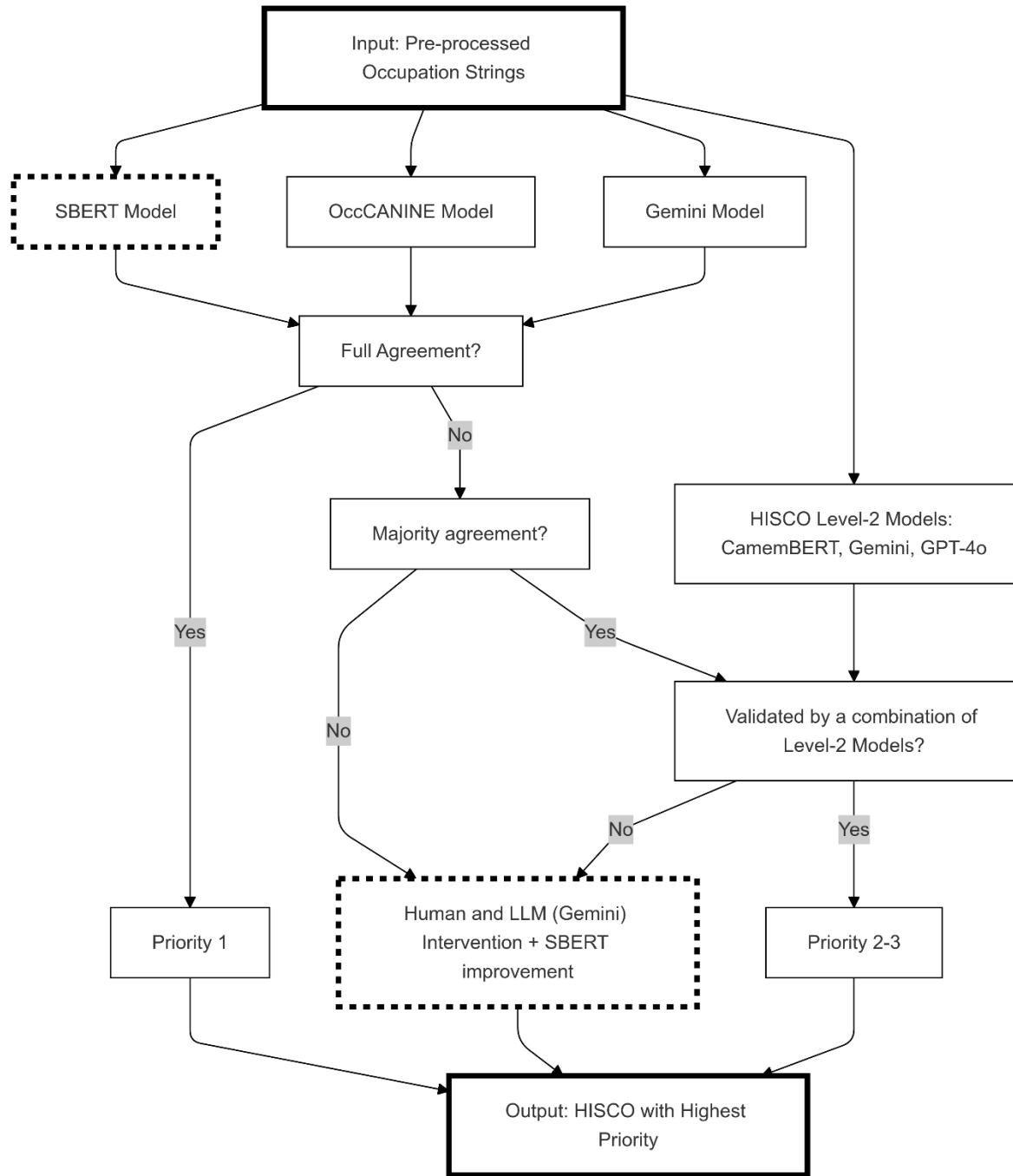
predict

Agree? —————→ then —————→ profit?

Large Language
Model B

predict

Diagram of our HISCO process using this Disagreement Modelling.



Thank You!

