



# MINISTÈRE DE LA JUSTICE

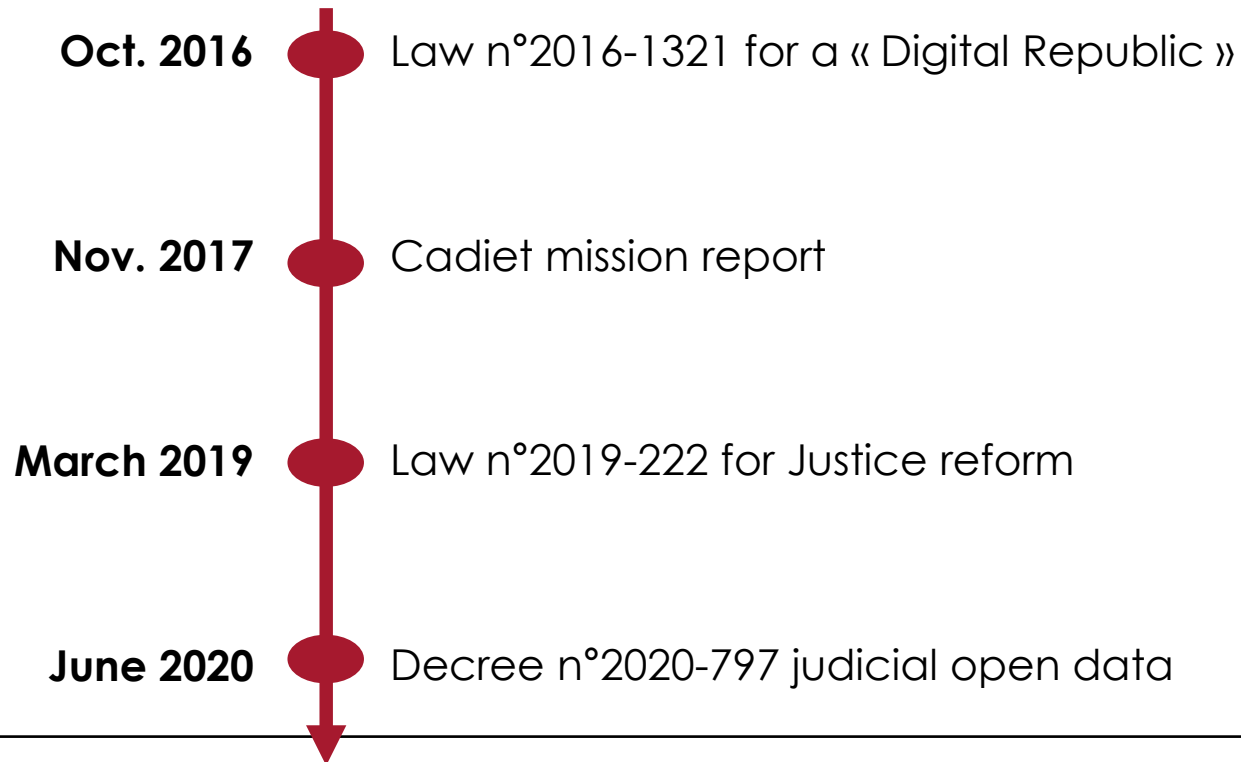
*Liberté*  
*Égalité*  
*Fraternité*

**The use of AI in the justice field -  
Anonymisation and pseudonymisation  
of judicial decisions**

**FRANCE**

# Evolution of the legal framework

## Judicial open data



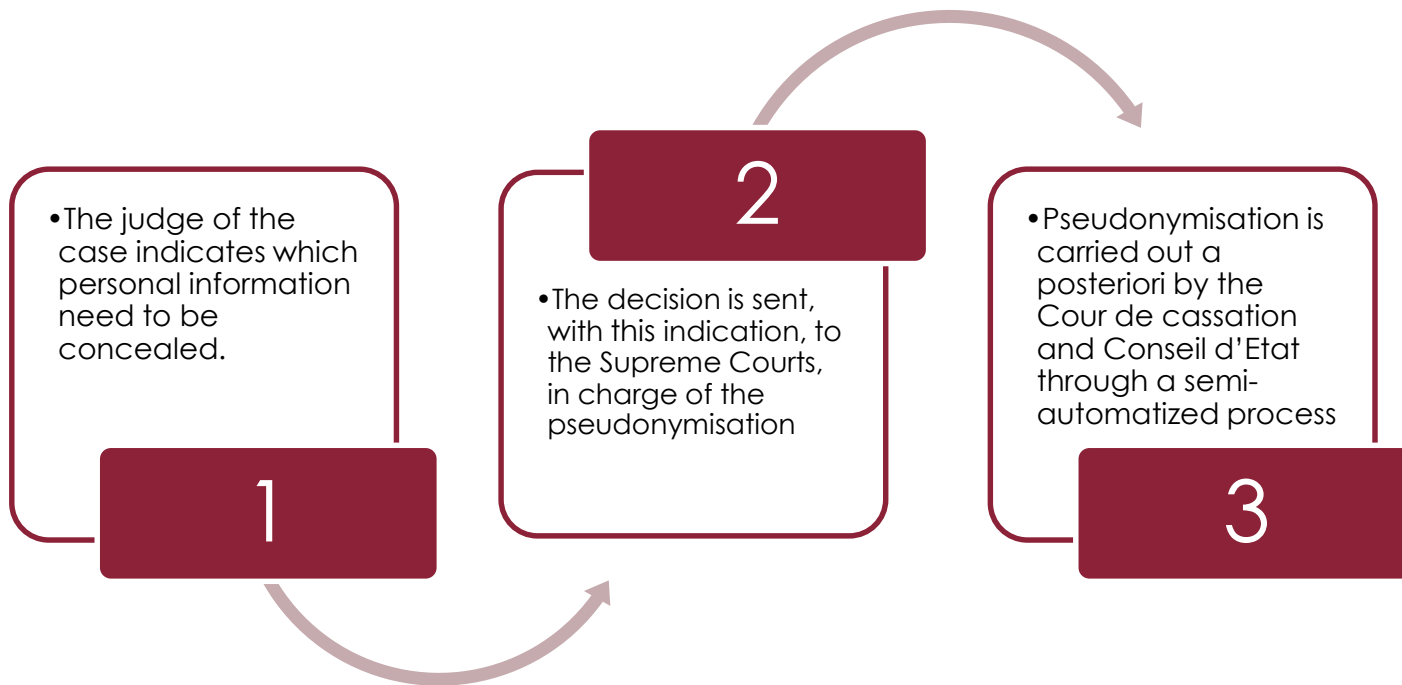
## Two pseudonymisation levels

The Open Data of court decisions must reconcile the objective of **transparency of justice** with the **protection of personal data**.

The 2020 decree specifies the rules for concealing the identification details of the persons mentioned in the decisions:

- **the surnames and forenames, systematically** ;
- and **additional concealment at the request of the judge** of the case, in the event of an infringement of their privacy or security.

# Pseudonymisation process « a posteriori »



# Technical developments

To implement this process, technical developments were necessary:

- evolution of applications** to allow judges to easily identify concealments to be made;
- creation of a robust and secure infrastructure;**
- development of **pseudonymisation tools;**
- development of **tools to make data available (API).**

# The project stages

2019

- Development of a pseudonymization engine (Cour de cassation)

2020

- Development of an annotation interface (Cour de cassation)

2021

- **Launch of pseudonymisation process**
- **Launch of open data web interfaces**
- **Open data of Cour de cassation and Conseil d'Etat decisions**

2022

- Open data of court of appeal decisions: administrative, civil, social and commercial matters
- Open data of first degree decision: administrative matters

A  
venir

- Open data of first degree decisions: civil, social and commercial matters
- Open data of criminal decisions

# Openness and dialogue

**Working groups aiming at defining concealment guidelines** to facilitate the work of judges and ensure relative harmonization, led by the Cour de cassation and Conseil d'Etat.

**Multidisciplinary working group on the reuse of data**, bringing together lawyers, legaltechs, publishers, researchers, to discuss and collect the expectations of these actors on technical issues, relating to the license or on ethical and regulatory subjects.

New tools of the Cour de cassation (pseudonymization engine, annotation interface, API) are **developed in open source and common standards**.



COUR DE CASSATION

Open data

Décisions de justice



## CHALLENGE : PROTECTION OF PERSONAL DATA



Court rulings have to be **pseudonymised prior to their dissemination**.



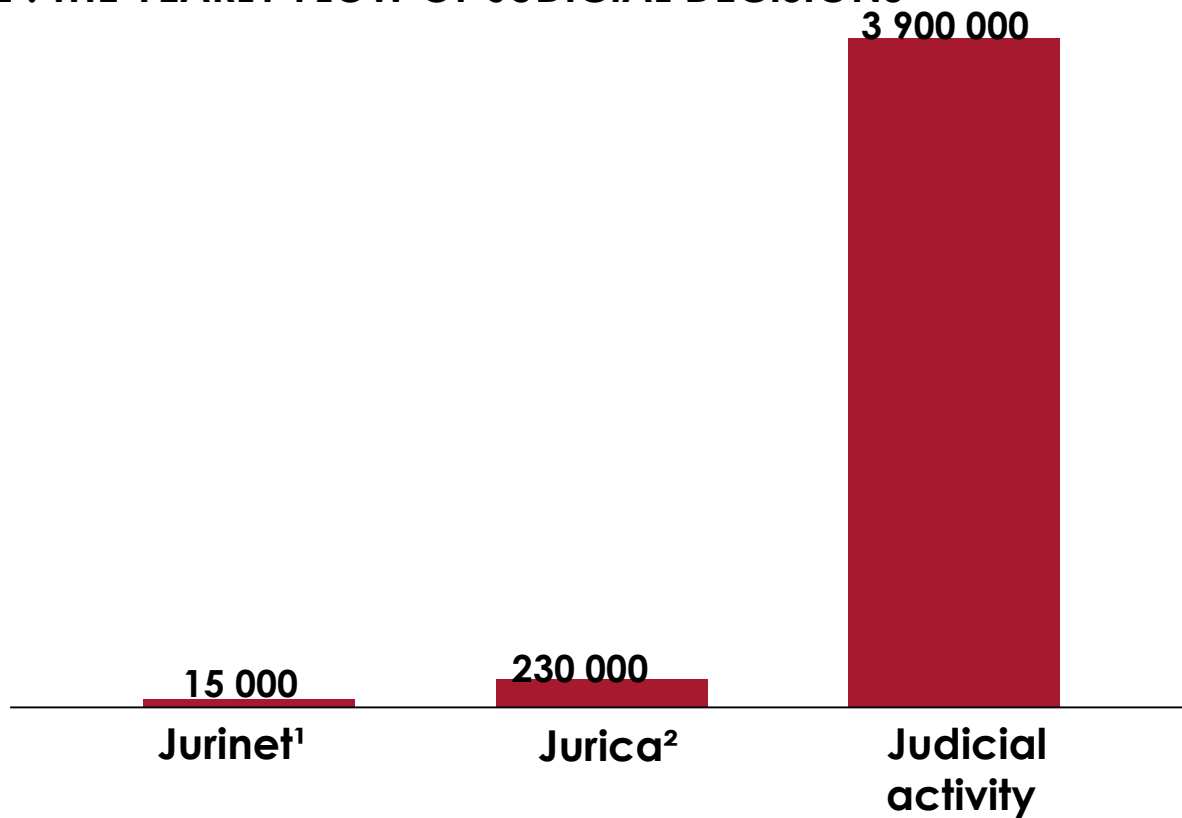
**Two pseudonymisation levels:** a « bedrock » covering of first names and surnames and complementary coverings decided by the judge to whom the dispute was attributed (privacy and security protection)



Easing the implementation : **formulation of harmonised doctrines of covering** by the creation of internal workshops (Cour de cassation and courts of appeal) – according to the **CNIL deliberation of February 6<sup>th</sup>, 2020**

---

## CHALLENGE : THE YEARLY FLOW OF JUDICIAL DECISIONS



<sup>1</sup> Cour de cassation rulings

<sup>2</sup> Courts of appeal rulings

## MACHINE LEARNING ENABLED PSEUDONYMISATION

Onboarding of **data scientists** in January 2019  
thanks to the **general interest entrepreneur** (eig3) programme.



# The pseudonymisation is a **named entity recognition** task

LA COUR DE CASSATION, PREMIÈRE CHAMBRE CIVILE, a rendu l'arrêt suivant :

Statuant sur le pourvoi formé par M. **Valentin PERSONNEPHYSIQUEPRENOM** **Barrière PERSONNEPHYSIQUEPRENOM** , domicilié **16 boulevard du Parc, 92200 Neuilly-sur-Seine ADRESSE** ,


contre l'arrêt rendu le 2 mars 2017 par la cour d'appel de Limoges (chambre civile), dans le litige l'opposant :

1°/ à Mme **Kim PERSONNEPHYSIQUEPRENOM** **Montalibet PERSONNEPHYSIQUEPRENOM** , épouse **Soriano PERSONNEPHYSIQUEPRENOM** , domiciliée **16 chemin de Saleve, 12130 Petit Lancy (Suisse) ADRESSE** ),

2°/ à M. **Amaury PERSONNEPHYSIQUEPRENOM** **Fouret PERSONNEPHYSIQUEPRENOM** , domicilié **2 rue des Piverts, 34000 Montpellier ADRESSE** ,

3°/ à M. **Cédric PERSONNEPHYSIQUEPRENOM** **Malherbe PERSONNEPHYSIQUEPRENOM** , domicilié **8 rue François Clouet, 76600 Le Havre ADRESSE** ,

4°/ à M. **Mathieu PERSONNEPHYSIQUEPRENOM** **Perez PERSONNEPHYSIQUEPRENOM** , domicilié **13 rue des Portes de Chanac, 19000 Tulle ADRESSE** .

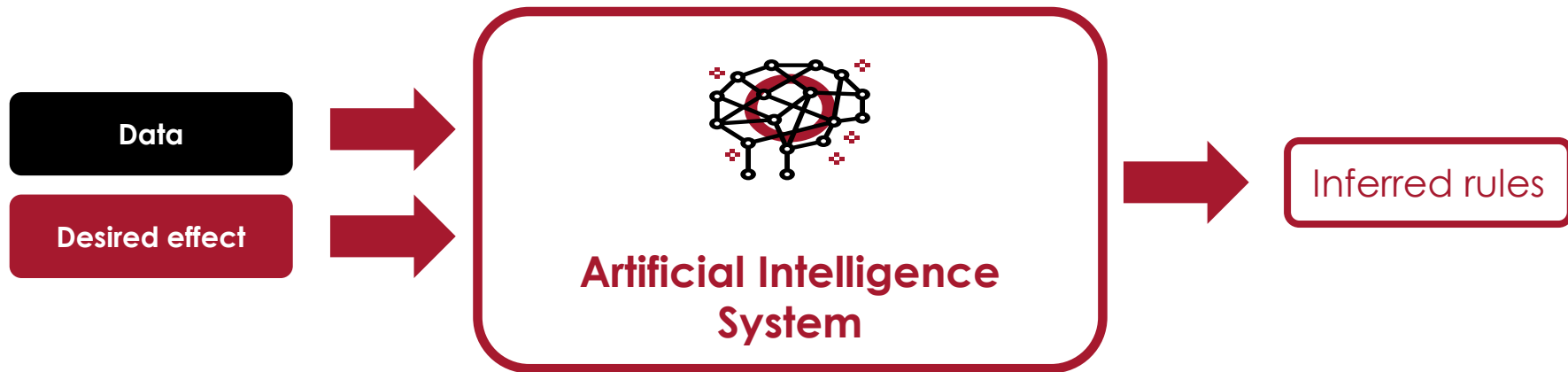


The entity « **Cédric** » belongs to a category  
« **personne physique prenom** » (first name of a  
natural person)

## ENTITES DETECTABLE BY THE MACHINE LEARNING MODEL

- Surnames and first names of natural person (parties or third parties)
  - Surnames and first names of professional (judge, clerk, lawyers)
  - Different civil dates (birth, death and marriage)
  - Addresses
  - Moral person
  - The registry
  - License plates
  - Bank accounts
  - The INSEE identifier number
  - Telephone / fax / mail number
-

## WHAT WE DO NOW



The system will learn all of the **inferred rules** to achieve the desired behavior.

To continue on the example of surname detection, the model will comprehensively integrate the different contexts in which a surname can appear.

---

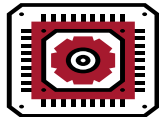
## REQUIREMENTS FOR AN AI PROJECT



Corpus of justice decisions



Annotated documents (Annotators)



Computing power(gpu)

---

## THE PERFORMANCE OF THE ML MODEL

### BEFORE

#### Rules engine Luxid

Surname natural person **85%**

### AFTER

#### Deep learning engine

Surnamme & first name  
natural person **98%**

Address **89,4%**

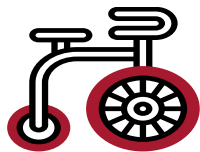
Dates **89%**

Moral person **92,8%**

**MANUAL CORRECTION TIME REDUCED BY 50%**



## GENERAL INTEREST PROGRAMMES



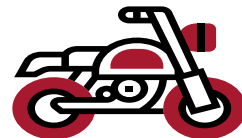
### EXTERNAL PROVIDER

Deterministic system,  
hard to improve.



### EIG 3

Artificial intelligence  
pseudonymisation  
system.



### EIG 4

Innovative human-  
machine annotation  
interface.

---

# THE CURRENT ANNOTATION AT THE COURT OF CASSATION

Systematic human verification, obsolete interface

Interface d'annotation

Document : JNT\_Cour de cassation\_Chambre civile 2\_20190321\_21900402\_1693391

Eléments annotés :

Eléments remplacés

**Personne physique (nom)**

	M	22	
	H	1	
	G	20	

**Personne physique (prénom)**

	K	1	
	V	1	
	L	6	

**Adresse**

	1	
	5	
	1	

Document d'origine

Document anonymisé

1  
2 CIV. 2  
3  
4 LG  
5  
6  
7  
8 COUR DE CASSATION  
9  
10  
11  
12 Audience publique du 21 mars 2019  
13  
14  
15  
16  
17 Rejet  
18  
19  
20 Mme FLISE, président

1  
2 CIV. 2  
3  
4 LG  
5  
6  
7  
8 COUR DE CASSATION  
9  
10  
11  
12 Audience publique du 21 mars 2019  
13  
14  
15  
16  
17 Rejet  
18  
19  
20 Mme FLISE, président

## MULTIPLES CATALYSTS



An analysis of the human annotations for continuous improvement.



Flow management



An ergonomic and user-friendly interface.



Clarification and simplification of the annotation rules.

---

# THE NEW INTERFACE (WORK IN PROGRESS)

Décision n°21000009 · Chambre civile · 07/01/2010



## Annotations demandées

Vue anonymisée

- Nom de famille (3)** ^
  - AUBERT 17
    - Aubert
  - Rogean 4
  - Windenberger-Jenner 4
- Prénom (3)** ^
- Adresse (6)** ^
- Personne morale (4)** ^
- Professionnel (nom) (11)** ^
- Localité (0)**
- Personne physico-morale (0)**
- Personne géo-morale (0)**
- Établissement (0)**
- Professionnel (prénom) (0)**
- Date de naissance (0)**
- Date de mariage (0)**
- Date de décès (0)**
- Réinitialiser**

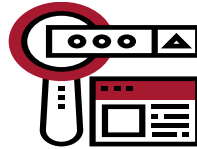
22  
23  
24 Arrêt n° 9 F-D  
25  
26 Pourvoi n° G 07-16.986  
27  
28  
29  
30  
31  
32  
33  
34 REPUBLIQUE FRANCAISE  
35  
36  
37  
38 AU NOM DU PEUPLE FRANCAIS  
39  
40  
41  
42 LA COUR DE CASSATION, DEUXIÈME CHAMBRE CIVILE, a rendu l'arrêt suivant :  
43  
44 Statuant sur le pourvoi formé par M. Jean-Yves Aubert, domicilié Maison d'arrêt de  
45 Varennes-le-Grand, n° 11463, A 302, BP 13, 71240 Varennes-le-Grand,  
46 contre l'arrêt rendu le 17 avril 2007 par la cour d'appel de Lyon (8e chambre  
47 civile), dans le litige l'opposant :  
48 1°/ à la Caisse de garantie des administrateurs judiciaires et des mandataires  
49 judiciaires à la liquidation des entreprises, dont le siège est 13 rue des  
50 Pyramides, 75001 Paris,  
51  
52 2°/ à Mme Sabienne Windenberger-Jenner, domiciliée 5 rue des Frères Lumières, Parc  
53 d'activités Eckbolsheim, 67087 Strasbourg cedex 2,  
54  
55

Valider

## ENHANCED DISSEMINATION OF CASE LAW



**New website** for the Cour de cassation



Efficient case law **search engine**

---



## Pseudonymisation of public decisions at Conseil d'Etat

29<sup>th</sup> of march 2021

## GENERAL OVERVIEW



- ❑ Pseudonymisation is applied on first names and surnames of physical persons
- ❑ To do so, the Conseil d'État owns two in-house softwares :
  - Determinist : developed on the basis of functional rules. In use since January 2013
  - IA /NLP : development on going
- ❑ Every day decisions are deposited on documentary bases : ArianeWeb and Légifrance by an automated EAI (enterprise application interface) process
- ❑ Decisions from Conseil d'État are published with their European Case Law Identifier (ECLI)

## DETERMINIST SOFTWARE



### ❑ Two components :

- An engine anonymising in backend all the decisions automatically sent by the EAI or manually sent by the users
- A Web interface to consult the decisions and to remove doubts by the curator



# DETERMINIST SOFTWARE



## Lots

Filtres

Type:  Manuel  Auto  Test

Date soumission du  au

Date de lancement du  au

Date fin anonymisation du  au

Juridictions:

- Toutes
- TRIBUNAL ADMINISTRATIF
- COUR ADMINISTRATIVE D'APPEL
- CONSEIL D'ÉTAT
- TRIBUNAL DES CONFLITS
- COUR NATIONALE DU DROIT D'ASILE
- COUR ADMINISTRATIVE D'APPEL DE PARIS

Statut lot:

Statut des documents:

N° de décision:

[RAZ des filtres](#) [Filtrer](#)

[Lancer](#) [Arrêter](#)

[+Ajout](#)

LOTS									
<input type="checkbox"/>	Jurisdiction	Type	Nb documents	Soumission	Lancement	Fin anonymisation	Statut	Résultat	Action
<input type="checkbox"/>	BGSA	Auto	1	24/03/2021 09:15:38	24/03/2021 09:17:47	24/03/2021 09:17:51	Finalisé	1 ok	<a href="#">Détail</a>
<input type="checkbox"/>	BGSA	Auto	13	23/03/2021 05:54:18	23/03/2021 05:56:33	23/03/2021 05:57:03	Finalisé	11 ok 2 corrigé	<a href="#">Détail</a>
<input type="checkbox"/>	8 ème chambre	Auto	20	23/03/2021 05:54:18	23/03/2021 05:56:00	23/03/2021 05:56:33	Finalisé	20 ok	<a href="#">Détail</a>
<input type="checkbox"/>	8 ème chambre	Auto	30	20/03/2021 06:46:23	20/03/2021 06:48:51	20/03/2021 06:49:43	Finalisé	28 ok 2 corrigé	<a href="#">Détail</a>
<input type="checkbox"/>	6 ème chambre	Auto	7	20/03/2021 06:46:23	20/03/2021 06:47:11	20/03/2021 06:47:25	Finalisé	6 ok 1 corrigé	<a href="#">Détail</a>
<input type="checkbox"/>	BGSA	Auto	1	19/03/2021 05:52:03	19/03/2021 05:53:35	19/03/2021 05:53:39	Finalisé	1 ok	<a href="#">Détail</a>
<input type="checkbox"/>	Référés	Auto	1	18/03/2021 06:12:46	18/03/2021 06:14:46	18/03/2021 06:15:04	Finalisé	1 corrigé	<a href="#">Détail</a>
<input type="checkbox"/>	BGSA	Auto	7	18/03/2021 06:12:46	18/03/2021 06:13:51	18/03/2021 06:14:06	Finalisé	7 ok	<a href="#">Détail</a>

Page 1 sur 1

Enregistrements 1 - 8 sur

[Lancer](#) [Arrêter](#)

[+Ajout](#)



## IA/NLP SOFTWARE



- A proof of concept has been developed
- We are working with external experts in data science and NLP to compare the two softwares and to train the algorithms
- Experimental mode