When AI goes fishing: On the importance of complete and representative training of AI-driven competition law enforcement

Jerome De Cooman – jerome.decooman@uliege.be

Workshop @ Comète on Ethical AI

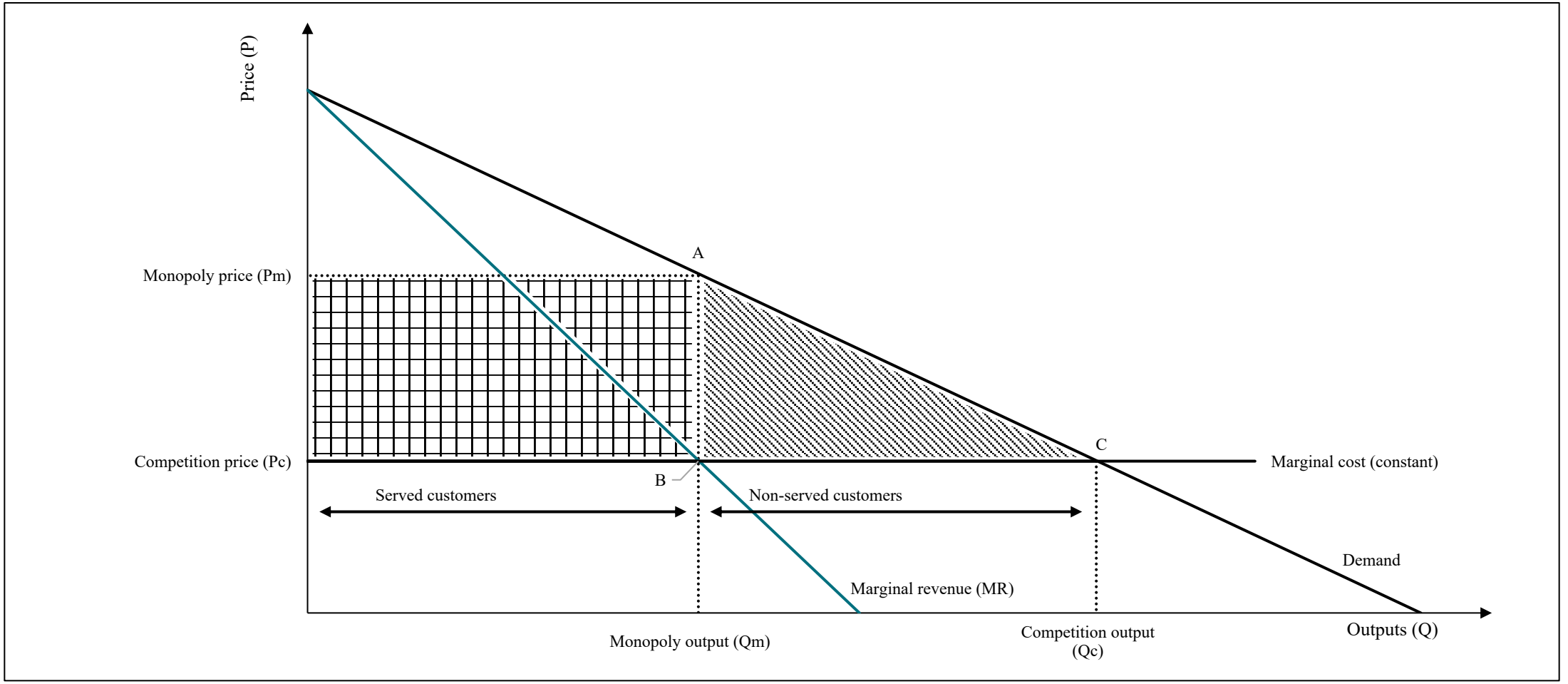
Inria, Ecole Polytechnique de Paris

23-24 November 2023 (Palaiseau, Paris, France)

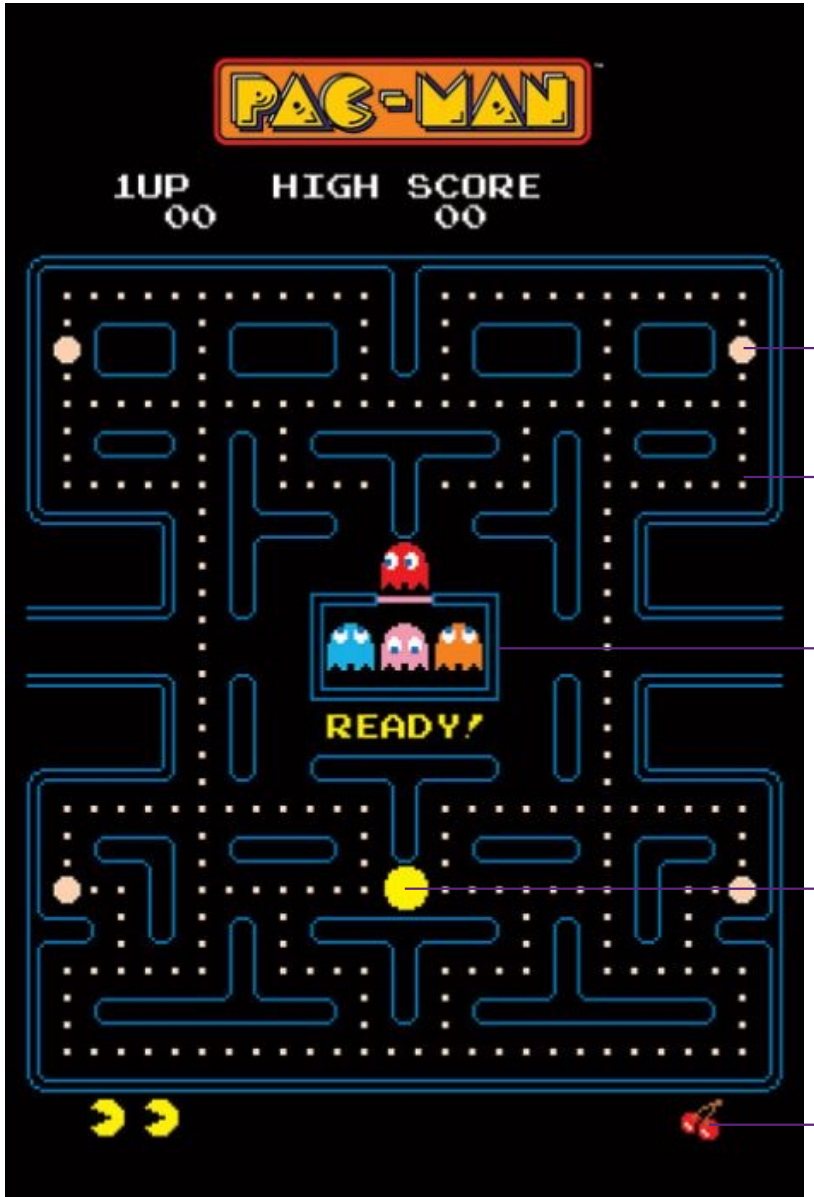
LIÈGE université
Cité

I. Introduction

I. Introduction: Standard monopoly model



I. Introduction: Pac-Man is a cartel (Schinkel 2014)



Leniency programme

Consumer surplus

(National) competition authorities

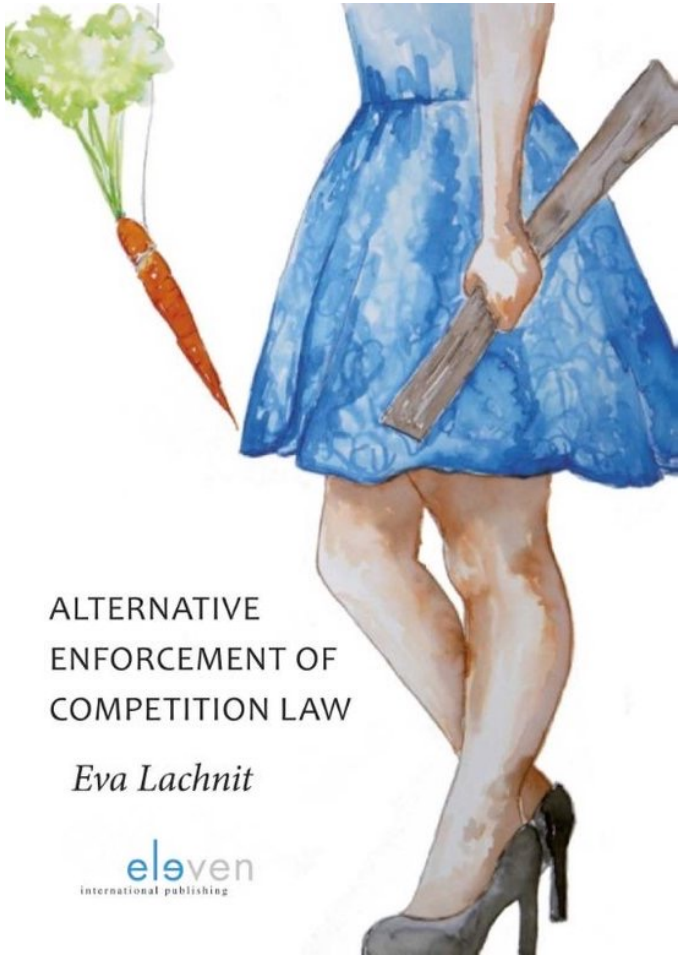
International cartel

Occasional windfall profit

LIÈGE université
Cité

II. Soft Carrots and Hard Sticks

II. Soft Carrots and Hard Sticks



- Undertakings are (supposed to be) rational utility maximisers
  - Participation constraint
    - cartelisation gain > sanction
  - Incentive constraint
    - cartelisation gain > deviating gain
- Two ways to destabilise a cartel
  - Hard sticks: Increasing sanction
  - Soft carrot: Increasing the gain driven from deviation

A. Hard Stick

- "The value of the punishment must not be less in any case than what is sufficient to outweigh that of the profit of the offence" (Bentham)
- Concretely:

$$E[S] > E[G]$$

A. Hard Stick

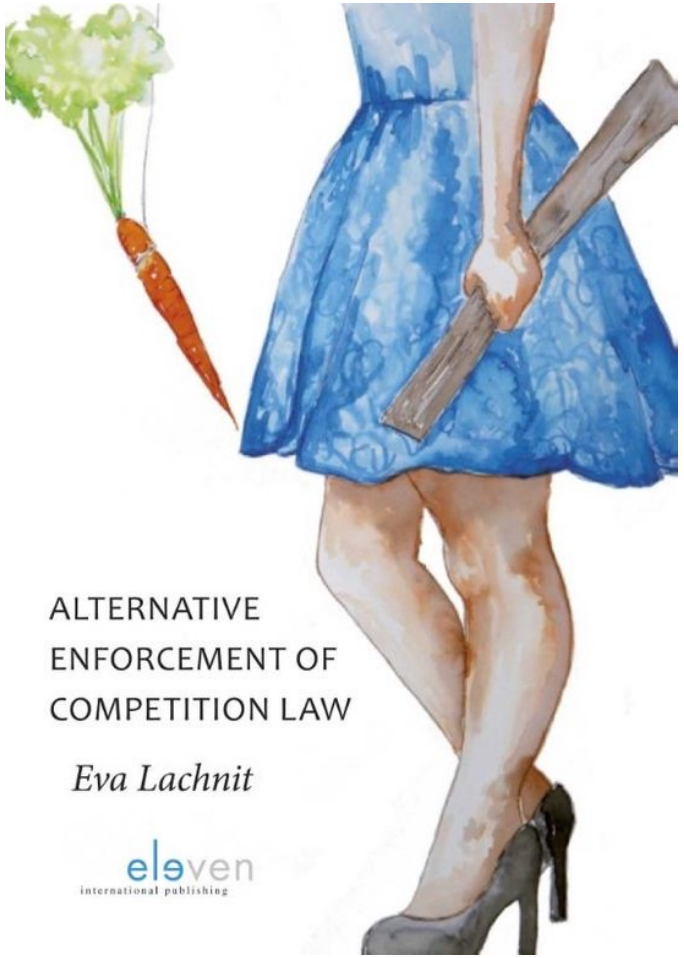
$E[S] > E[G]$	E[G] = 10 E[S] > 10
$E[S] > \frac{E[G]}{P_D}$	P_D = 0.1 E[S] > 10/0.1 = 100
$E[S] > \left(\frac{E[G]}{P_D} + C_{enforcement}\right)$	$C_{enforcement}$ = 5 E[S] > 10/0.1 +5 = 105
$E[S] > \left(\frac{E[G]}{P_D} + C_{enforcement}\right) * A$	A = 1 → risk neutral A < 1 → risk averse A > 1 → risk taker

B. Soft Carrots

- Leniency: increasing detection

Position	Reduction	Reduction after investigation starts
1st	Immunity	30-100%
2nd	30-50% of the fine	20-30%
3rd	20-30%	Up to 20%
Subsequent	Up to 20%	

II. Soft Carrots and Hard Sticks



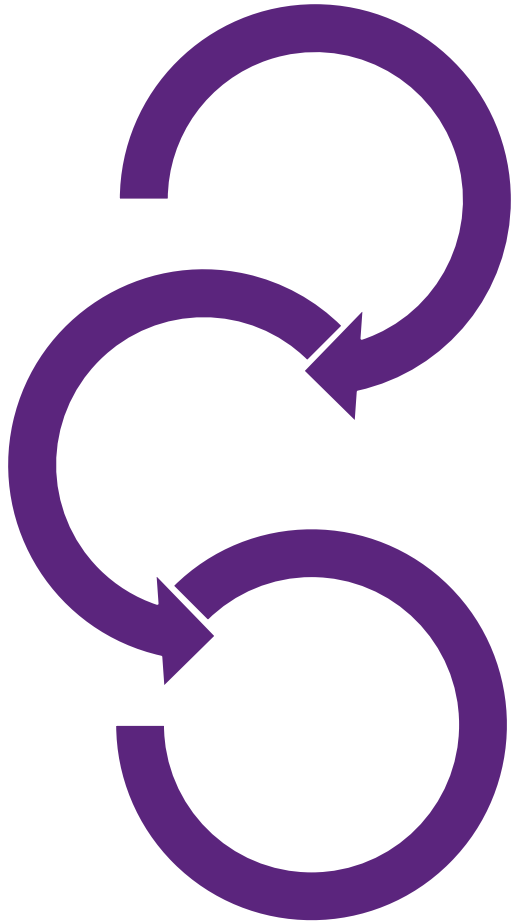
- Controversial success
  - Sanction is a function of detection
  - Yet, probability of detection is low
  - Leniency is a way to increase the probability of detection
  - Yet, probability of detection decreases the effectiveness of leniency
- "While there is a recognition that a leniency program is an immensely valuable tool (…) concerns arise when it is the *only* tool" (Harrington and Chang 2015)

LIÈGE université
Cité

III. Alternative: Cartel Screening

III. Algorithmic shift in the fight against cartels

- AI systems draw the sketch of suspicious businesses by identifying cartelists' recurring characteristics or patterns (Sanchez-Graells 2019)



Increasing the probability of detection…

…Increases the incentive to apply for leniency…

…that in turn deters cartel formation.

III. Alternative: cartel screening

- How does it works?
- There is "*conventional wisdom on collusion*" that permits the identification of "*factors that are supposed to hinder or facilitate*" collusive behaviours (Tirole 1988)
  - Structural screens: analysis of market structure
  - Behavioural screens: analysis of the collusive methods or outcome of collusion

Structural screens

Structural screens		High probability of cartelisation
Structural factors	Number of firms (concentration)	Low (high)
	Entry barriers	High
	Undertakings' interaction	Frequent
	Transparency	Low demand side, high supply side
Supply-side factors	Vertical product differentiation	Homogeneous product
	Innovation	Low-innovative markets
	Advertisement	Low-advertising industries
Demand-side factors	Demand	Stable
	Buyer bargaining power	Low
	Horizontal product differentiation	Low differentiation

Behavioural screens

Collusive markers		Collusive behaviour
Price	Price evolution	Low variance Sharp increase in high price-cost margin Sharp decline of price followed by sharp increase
	Product price and quality	Homogenisation through increased product standardisation and pricing formula
	Prices across customers	Decrease of customer-specific prices
Market shares	Sales quotas	Distribution of market shares seems more stable under collusion
	Exclusive territories	Price increase in the home-market, export decreases
	Customer allocation	Stable customer base
Enforcement	Buy-back	In time *t* a firm A sells above its historical market share while a firm B sells below its historical market share; in *t+1*, A buys products from B
	Compensation	In time *t* a firm A sells above its historical market share while a firm B sells below its historical market share; in *t+1* the sale levels are inverted

III. Alternative: cartel screening

Cartel screening *does*

- Screens identify patterns of collusion
- Screens trigger the need for, *e.g.*, dawn raids

Cartel screening *does not*

- Screens do not prove collusion

III. Alternative: cartel screening

- Studies demonstrate (AI-driven) cartel screening works
  - Detection of illegal agreements (Coglianese and Lai 2022)
  - Detection of corruption (*e.g.*, in public procurement)
- However, (AI-driven) cartel screening "still has sceptics" (Abrantes-Metz 2014)
- This algorithmic solution faces three challenges (De Cooman 2023)
  - Data issue (availability of quality data)
  - Algorithmic issue (duty to state reasons when the weight of the parameters are unknown?)
  - Human issue (automation bias)

LIÈGE université
Cité

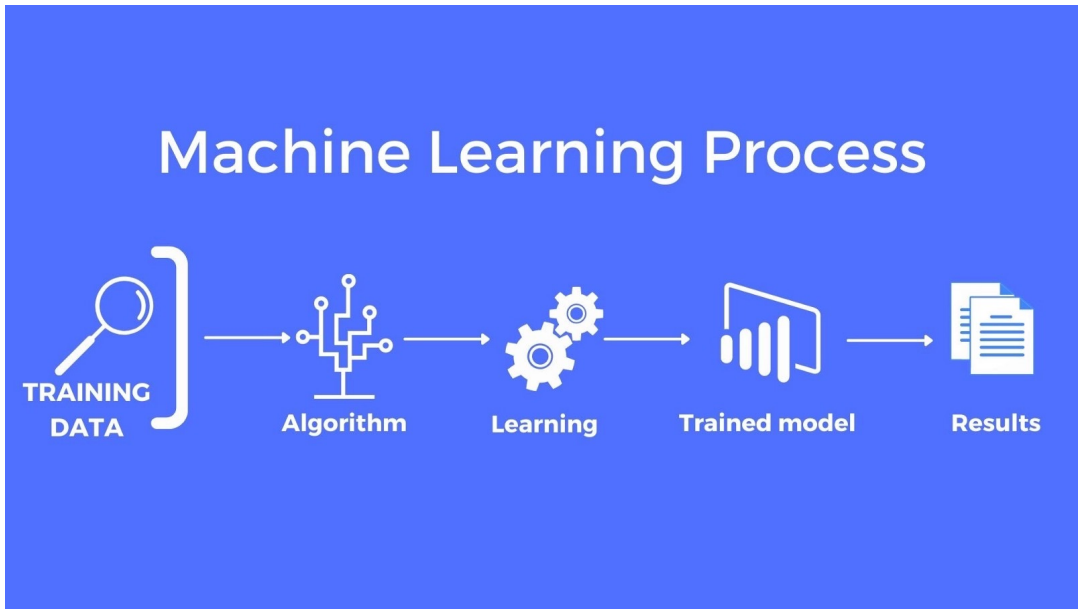
IV. Pitfalls

A. Data availability

B. Data quality

A. Data availability

"no data, no fun"



- Detect collusion in dataset T (for target)
- Trained on
  - Dataset W (same market) or;
  - Dataset Z (comparable market)
- No dataset W or Z?
  - No training or incomplete training
- *Caveat*: UK CMA (W = 100 tenders and 500 bids)

B. Data quality

“Dirty data, bad prediction”

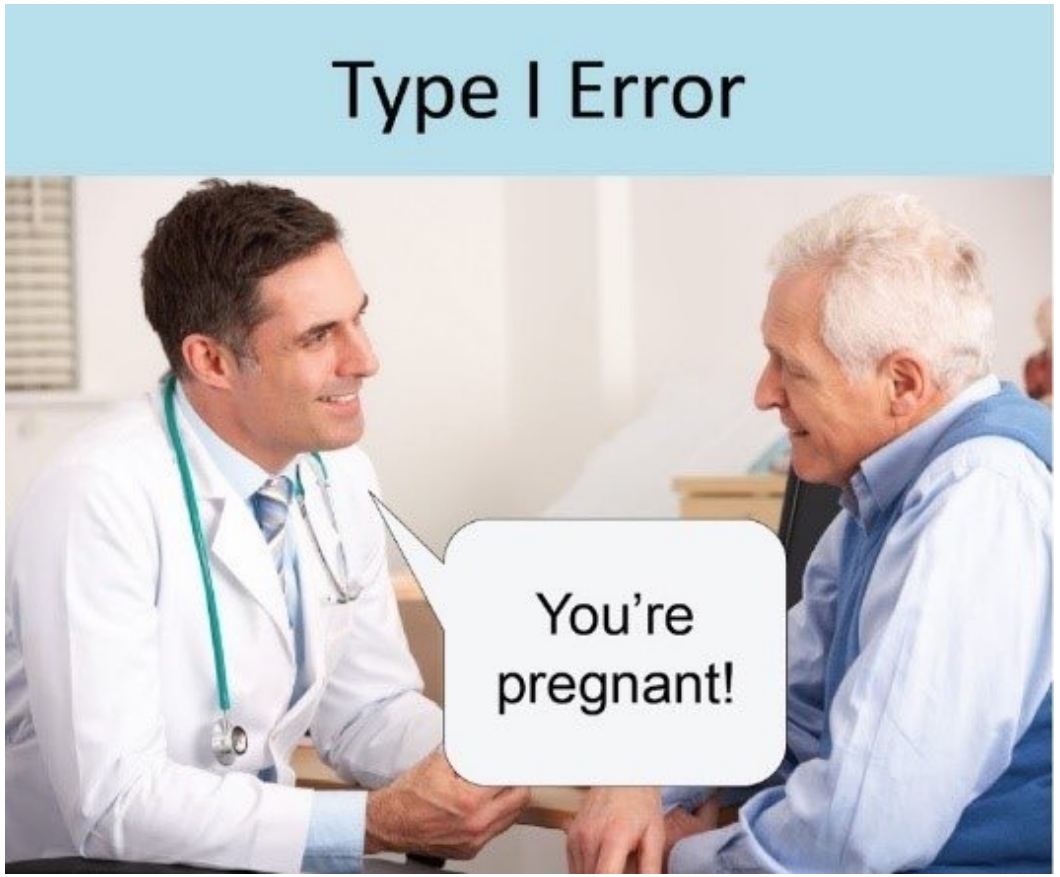
- An elephant tale
- Closed-game scenario
- Statistically representative data

1. Risk of Type II Error – non-detection of cartel

- "With oligopoly, everything is possible" (Stiegler 1964)
  - *Animal Feed Phosphate Cartel*
- Interaction between collusive markers
  - In principle, demand fluctuations hinder collusion
  - But an increase in demand fosters collusion when entry barriers are sufficiently high
- Selection bias
  - Are discovered cartel statistically representative of the whole population of cartel?

2. Risk of Type I error – Mistakenly detecting a cartel

- Erroneously investigate a competitive behaviour
- Waste of time and resources
  - Subsequent trust?
- Pricing Parallelism vs anticompetitive behaviour
  - Ultrasound and biopsy
  - Does it matter?



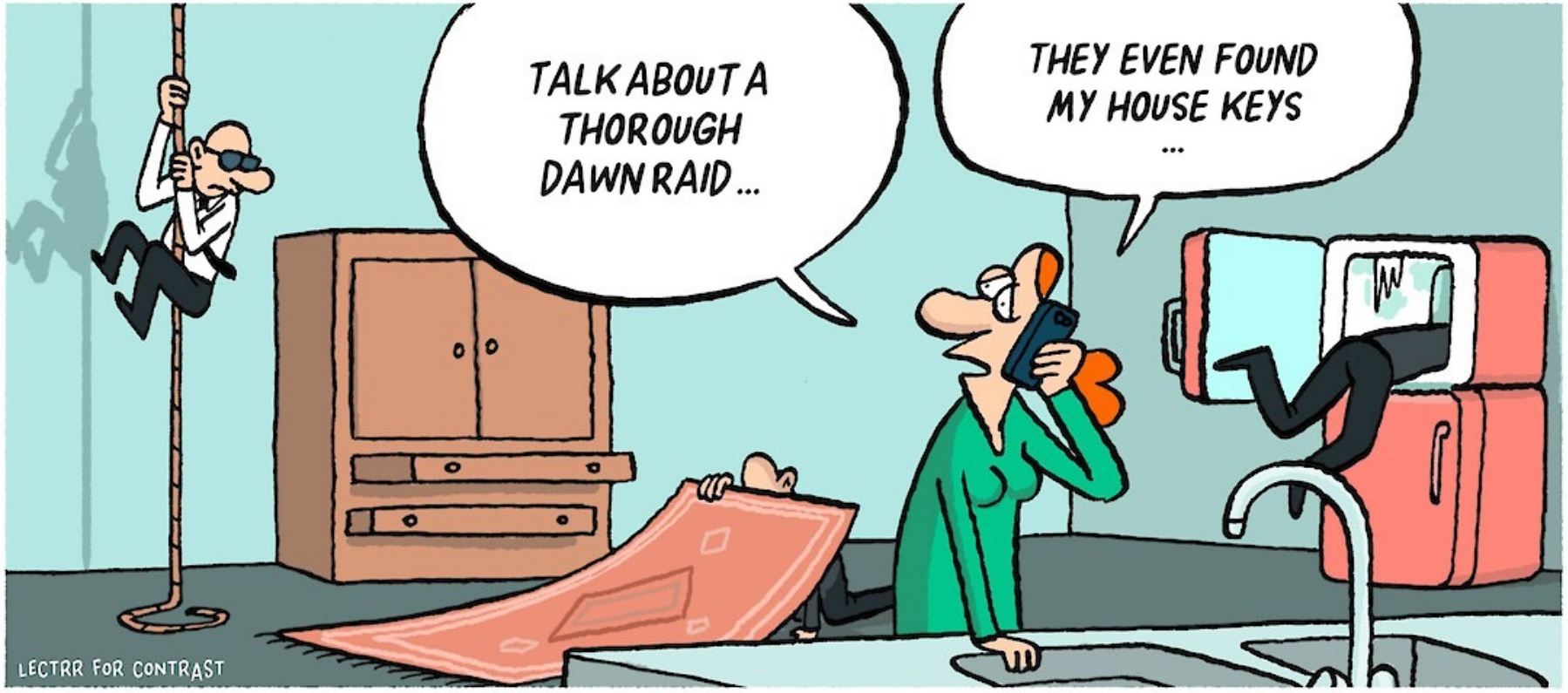
3. Why it does matter: dawn raid or fishing expedition?

- "Obligation to state specific reasons" (Art. 20(4) Reg. 1/2003)
  - Description of the features and nature of the suspected infringement
  - Presumed Facts the EC intends to investigate
  - Sectors and market thought to be affected
- A decision authorising a dawn raid written "in very general terms" might still contain "the essential indications prescribed by Art. 20(4) Reg. 1/2003" (*Dow Benelux* § 11)

Nexans (C-37/13) and Prysmian (T-140/09)

3. Why it does matter

- To be in possession of "information and evidence providing *reasonable grounds* for suspecting infringement of the competition rules by the undertaking concerned" (*Roquette Frères*, §61)



The EC "has not demonstrated that it had reasonable grounds for ordering an inspection covering all electric cables" (Nexans, §91; Prysmian, §89)

3. Why it does matter

- Is a red flag raised by cartel screening a reasonable grounds for suspicion? (*Roquette frères*)
- Is the statement of reasons "excessively succinct, vague and generic"? (*Heidelberger Cement*, § 39)
- It depends on the level of Type I Error

3. Why it does matter: a parallel with IP

- A filtering system with an *inadequate* rate of false positive would be contrary to fundamental rights (Case C-360/10, § 50)
- A filtering system "shall not result in the prevention of the availability of works or other subject matter uploaded by users, which do not infringe copyright and related rights" (art. 17(7) Dir. 2019/790)
- Shall not suggests a type I error-free filtering system
  - Not true: it only suggests a "as low as possible" rate (AG Henrik Saugmandsgaard Øe C-401/19)


iStock™
Credit: Julia Garan
1194927414

LIÈGE université
Cité

V. Data Governance

V. Data governance

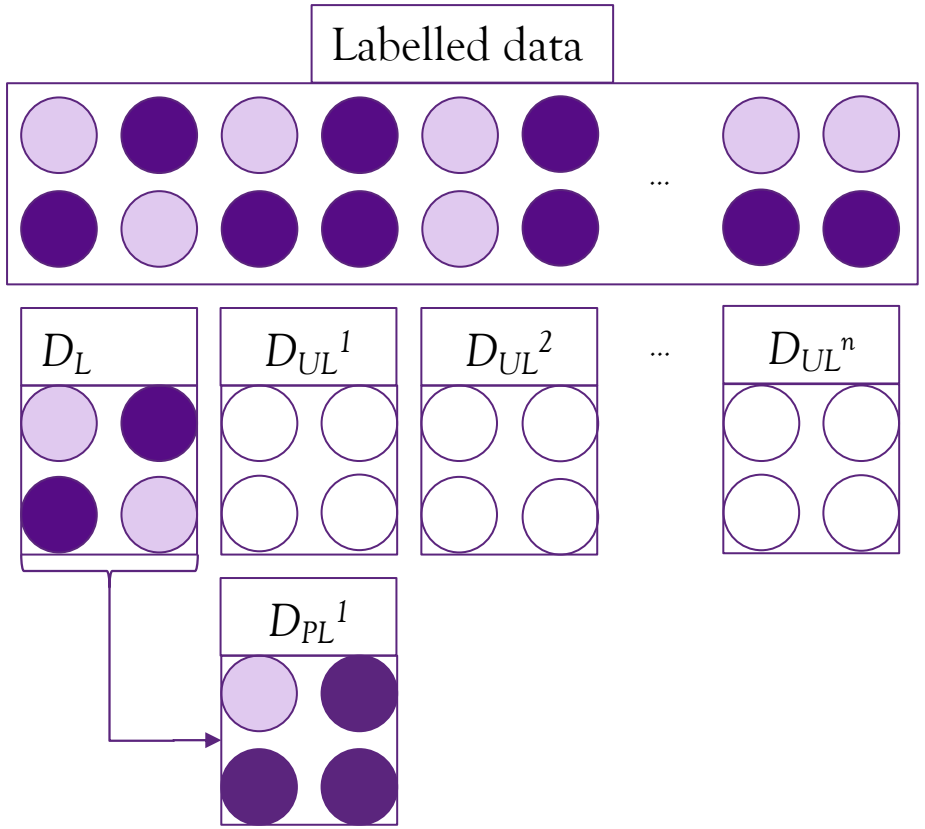
- Competition authorities do not want to lag behind (*e.g.*, UK CMA)
- "*It is a capital mistake to theorize before one has data*" (Conan Doyle 1889)
  - The algorithmic cart should not be put before the data horse
  - Construct a better data architecture before developing AI-driven cartel screening

V. Data governance

- Dataset composed of true positive and negative, false positive and negative
- Ideally, training on correctly labelled data
- If not, poisoning effect
- Relabelling has a cost (Sanchez-Graells 2021)

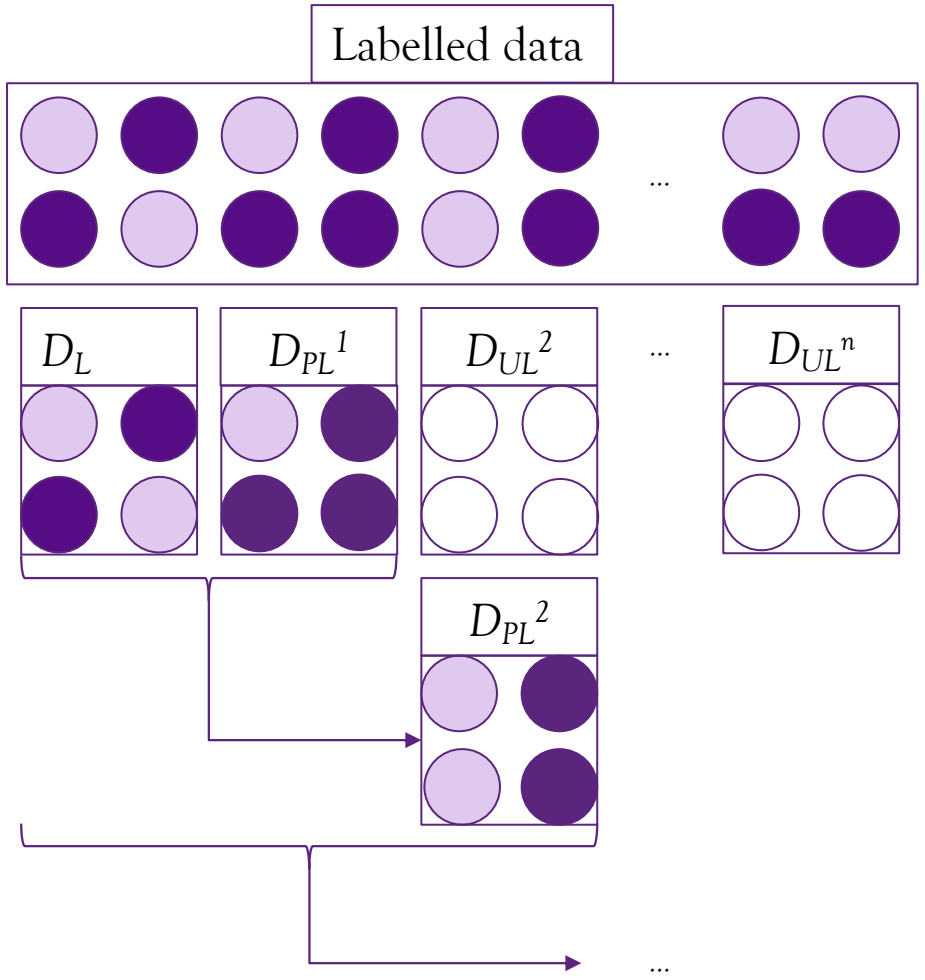
Éléphant blanc du roi de Siam.

V. Data governance



- Select small amount of labelled data D_L
- Unchecked data are de-labelled $D_{UL}^{1, 2, \ldots n}$
- Train the model on D_L
- Use the model to predict the labels of D_{UL}^1
- Unlabelled D_{UL}^1 become pseudo-labelled D_{PL}^1

III. Data governance



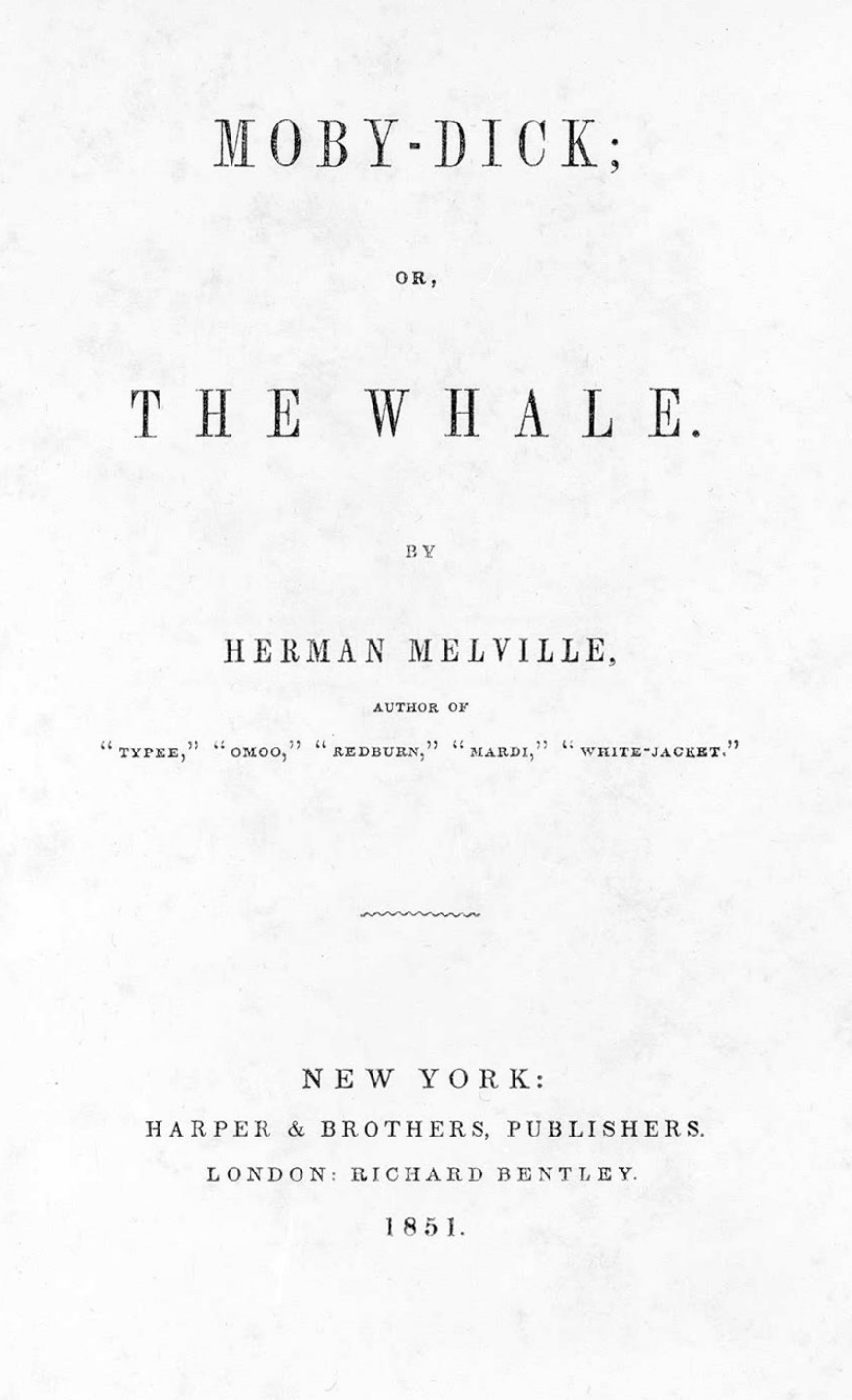
- Retrain the model on D_L + D_{PL}^1
- Use the model to predict the labels of D_{UL}^2
- D_{UL}^2 become D_{PL}^2
- Retrain the model on D_L + D_{PL}^1 + D_{PL}^2
- *Etc.*

LIÈGE université
Cité

VI. Conclusion

VI. Conclusion

- "*When all possibilities* (...) *become probabilities, every possibility is the next thing to a certainty*" (Melville, *Moby Dick*, 1851)
- Competition authorities have to remain aware that
  - AI-driven cartel screening increases effectiveness of EU law proceedings
  - It revitalizes *ex officio* investigations
  - It is a useful complement to leniency programme
- That said:
  - This requires available quality data


MOBY-DICK;

OR,

THE WHALE.

BY

HERMAN MELVILLE,

AUTHOR OF

"TYPEE," "OMOO," "REDBURN," "MARDI," "WHITE-JACKET."

NEW YORK:

HARPER & BROTHERS, PUBLISHERS.

LONDON: RICHARD BENTLEY.

1851.

LIÈGE université
Cité