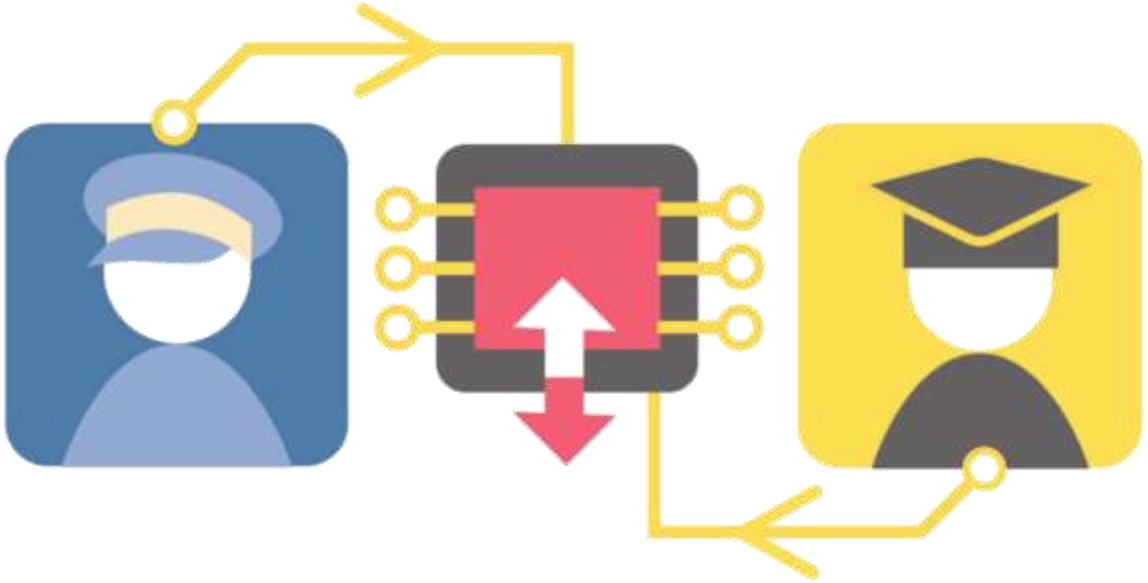
Utrecht University

9-10-2023

The National Police Lab AI

Researching, developing and evaluating AI for the Netherlands Police



Floris Bex

Scientific Director National Police Lab AI (Utrecht)
Associate Professor Artificial Intelligence (Utrecht University)
Full Professor Data Science and the Judiciary (Tilburg University)

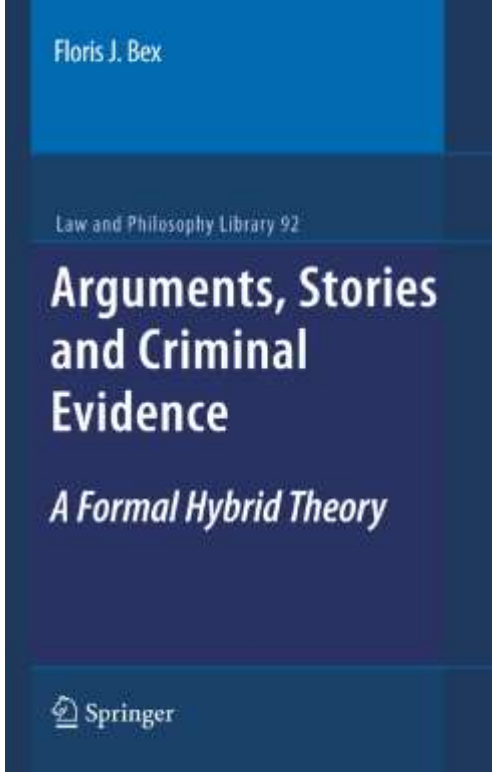
Floris Bex

- Full Professor Data Science & Judiciary (Law - Tilburg)
  - Together with Dutch Council for the Judiciary (Raad v.d. Rechtspraak)
  - AI for Law, Law for AI
- Associate Professor AI (Computer Science - Utrecht)
  - Argumentation in AI, Natural Language Processing, AI tools for forensic & legal reasoning

Evidence & AI

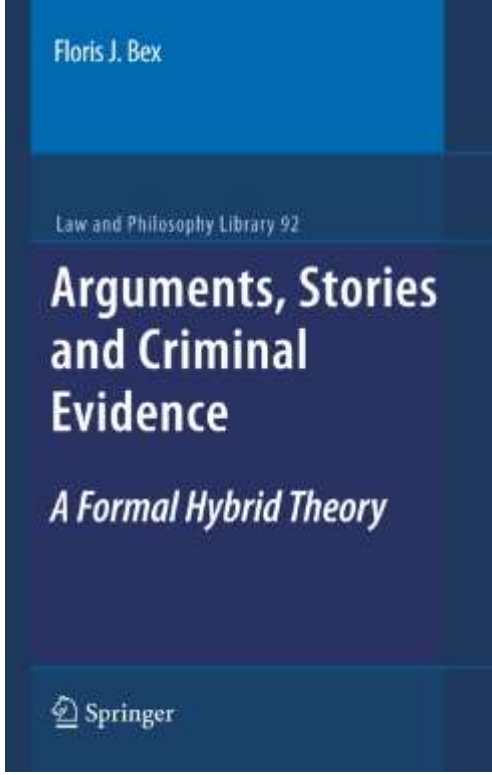
- The logic of criminal evidence
  - Factual arguments
  - Legal arguments

- Models of Rational Proof in Criminal Law
  - Henry Prakken, Floris Bex and Anne Ruth Mackor
  - Logical models, Bayesian models, Cognitive models

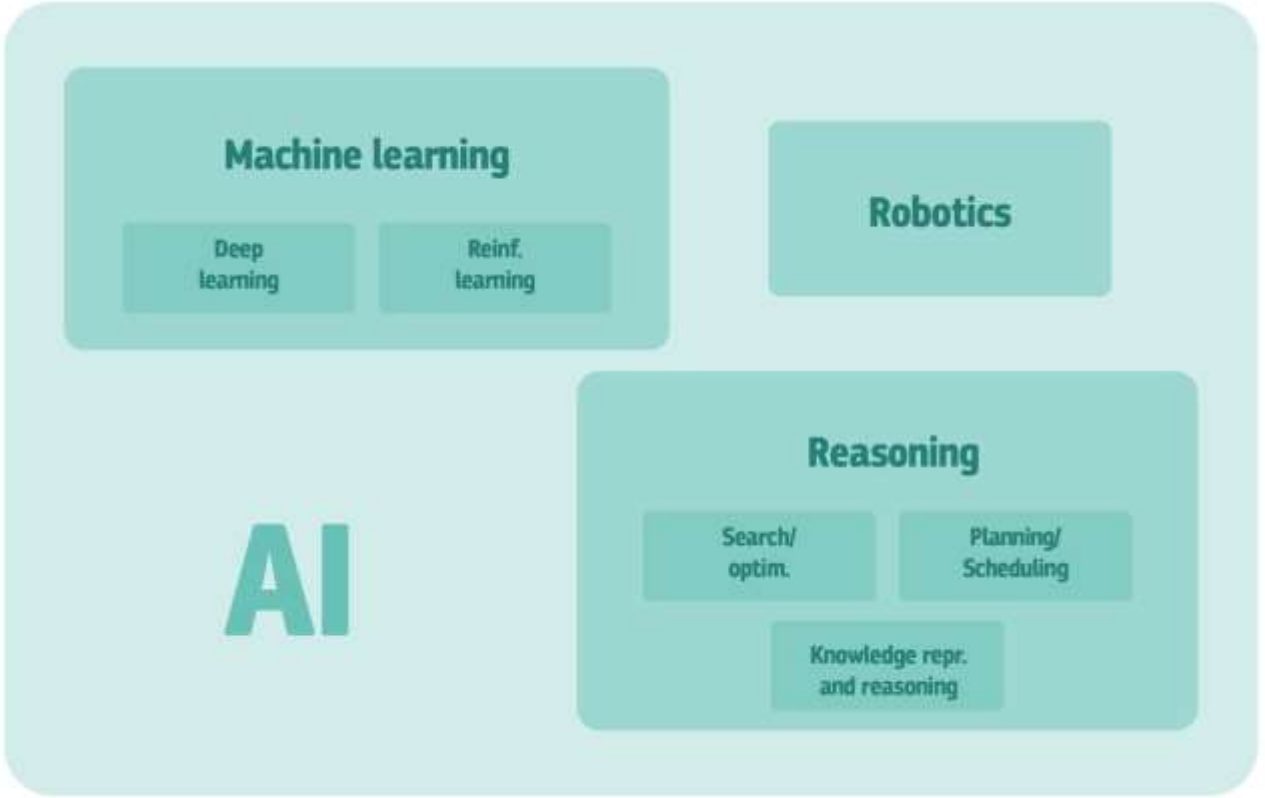


Data + knowledge for reasoning with evidence

- Data: observations from the environment
  - Observations by police officers, witnesses, camera's, algorithms
- Knowledge: argument types and scenario (story) types
  - Argument rules: argument from witness testimony, argument from statistics
  - Scenarios: fraud scenario, murder scenario, drug crime scenario

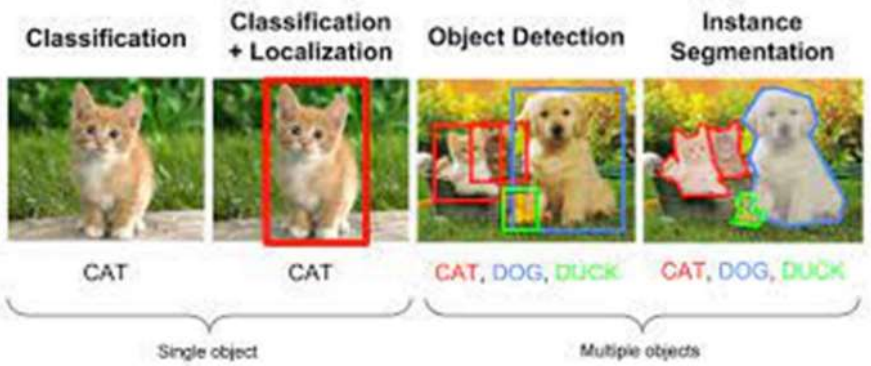


What is AI?



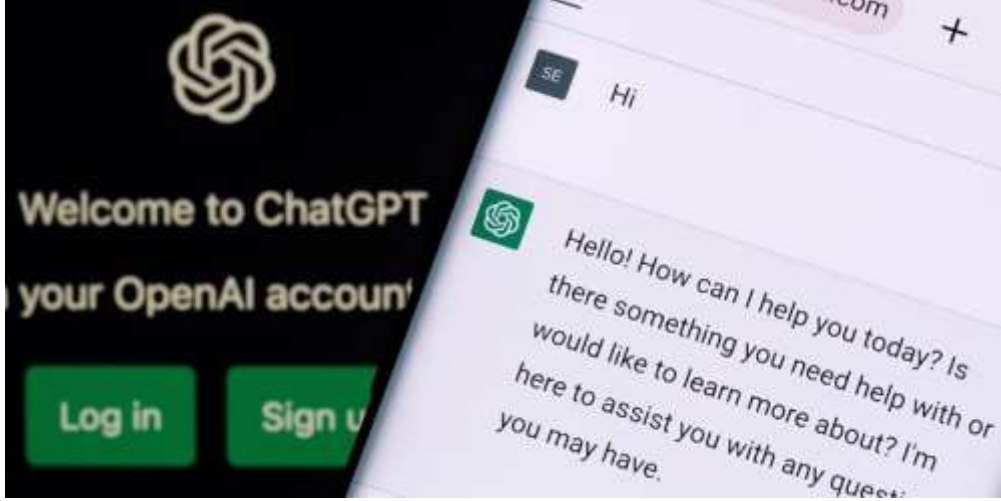
- Reasoning & learning
  - *HLEG AI definition*
- Machine learning & logic- and knowledge-based approaches
  - *draft AI act (Annex I)*
- Data-driven law & code/rule-driven law
  - *Hildebrandt et al.*
- (Data-driven) Machine Learning & Knowledge-based systems

Is this AI?



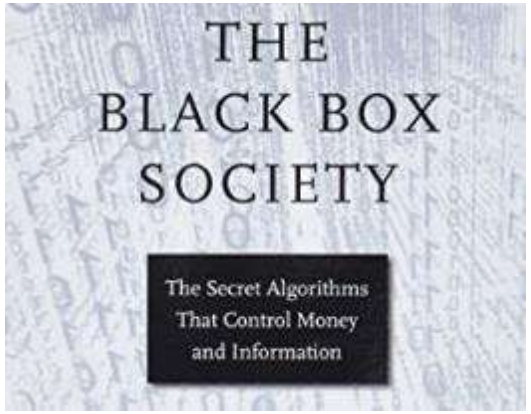
- Being able to distinguish cats from dogs?
- Win a game of chess? Win a game of Go?
- Determine how much tax someone has to pay?
- Driving a car in traffic? Controlling a vacuum cleaner the room?

Is this AI?



- Do your homework
- Writing and translating cover letters
- Writing computer code
- Create websites
- Summarize court rulings at B1 level
- ….

What is AI?
AI as a Rhetorical Tool



Frank Pasquale

We are caught up in a drama[1] between techno-sceptics and techno-optimists

Henrik Sætra

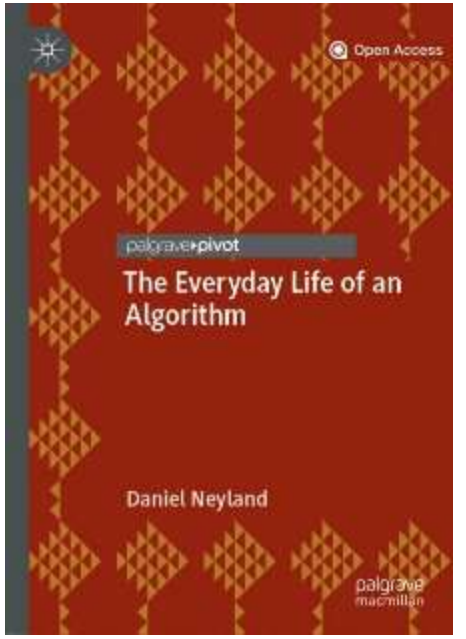
Regulating and governing AI

Building and applying AI

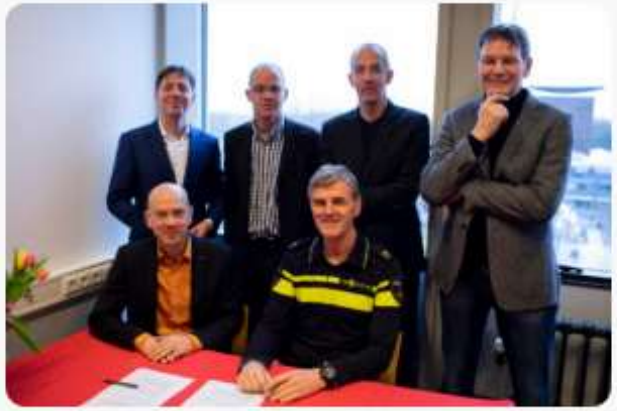
1. Ziewitz, M. (2016). Governing algorithms: Myth, mess, and methods. *Science, Technology, & Human Values*, *41*(1), 3-16. (Thanks to Daan Kolkman for pointing this one out!)

No more drama?

- Look at the everyday life of/with algorithms
  - Practical use cases of AI in context at the police
- Work across disciplines
  - Design, build, evaluate AI systems from different (disciplinary) perspectives
- The National Police lab AI 2019-2023
- Examples of AI at the Dutch Police (Lab)



The National Police Lab AI - 2019



- Collaboration of universities and police
- 9 computer science/AI PhDs
- Research & development of state-of-the-art AI for real police problems

uu.nl/en/research/ai-labs/national-policelab-ai

The National Police Lab AI - 2023

- Researchers from public management, media studies, law
- Not just build AI, but also evaluate it broadly
  - Sensitivity to public values, transparency, citizen trust, legal framework
- Research in collaboration with the lab
- 22 PhDs, about 1/3 with a non-CS/AI background

From Multimodal Data to Trustworthy Evidence in Court

AI4Intelligence

algosoc Justice

ALGOPOL: Research into Value-Sensitive and Transparent Algoritmization in the Police

AI for the Netherlands Police

- AI for (smart) search through digital evidence
- AI for supporting evidence gathering and analysis
- AI for citizen engagement
- AI for strategic analysis
- AI for automating routine cases.

Which type of AI?

- Machine Learning for (big, unstructured) data
  - Speech, text, images
  - When you want to search in a lot of data
  - When high-level conclusions are not necessary
- Reasoning for making/supporting decisions
  - Laws and regulations
  - Arguments and scenarios
  - When you want to automate tasks
  - When the data is structured and domain is restricted

Combating online trade fraud using hybrid AI (machine learning + reasoning)

Example 1: AI for citizen complaint/report intake

- Trade fraud: false webshops, malicious traders on Ebay
  - 40,000+ reports of alleged online fraud per year
  - Not all fraud: wrong product, not paid
- Automatically recommend to file report or not
  - Citizen fills in a form w. details & free text story
  - Possible fraud or not?

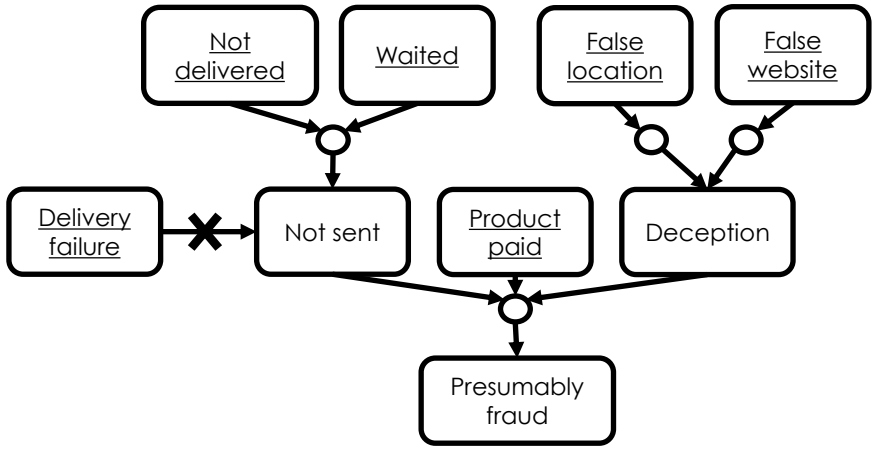
AI for intake – data & knowledge

- Combine data- and knowledge-driven AI
  - Relevant legal rules are known, bounded domain
  - Free-text interpretation needs data-driven AI

Schraagen, Testerink, Odekerken, Bex (2018)
Argumentation-driven information extraction
for online crime reports. *LeDAM 2018*

AI for intake – legal model

Legal model



Computational argumentation
Rules w. exceptions based on DCC & police policy rules

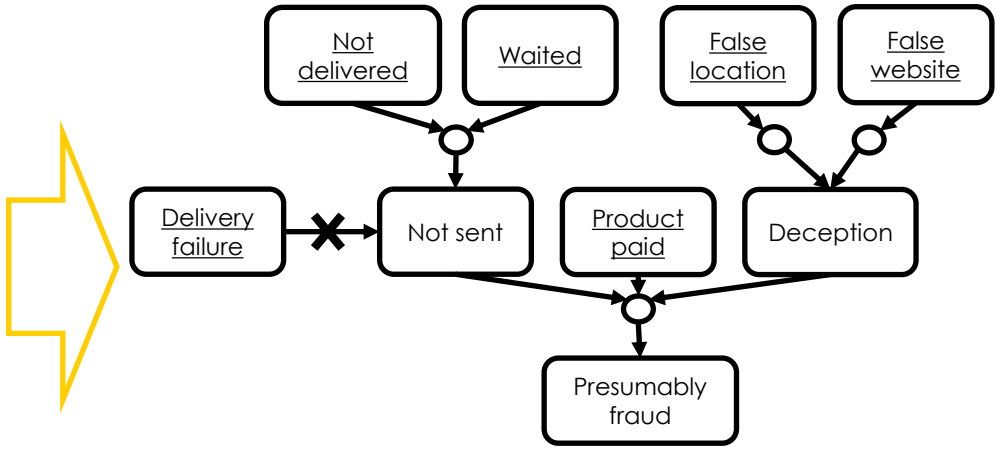
Schraagen, Brinkhuis & Bex (2017) Evaluation of Named Entity Recognition in Dutch Online Criminal Complaints. *DESI VII @ ICAIL 2017*

AI for intake – free text

Complaint form

Fictitious example report 1
I would like to report fraud. I recently saw a bicycle for sale on eBay and contacted the advertiser. He said he lived far away, so he would send me the bike. I paid him in good faith, but have still not received anything. I saw on Facebook he lives nearby.

Legal model



Computational argumentation
Rules w. exceptions based on DCC & police policy rules

D. Odekerken, F. Bex, A. Borg, B. Testerink (2022) Approximating Stability for Applied Argument-based Inquiry. *Intelligent Systems with Applications*.

AI for intake – combining IR and argumentation

Extracting observations from complaint form

Paid

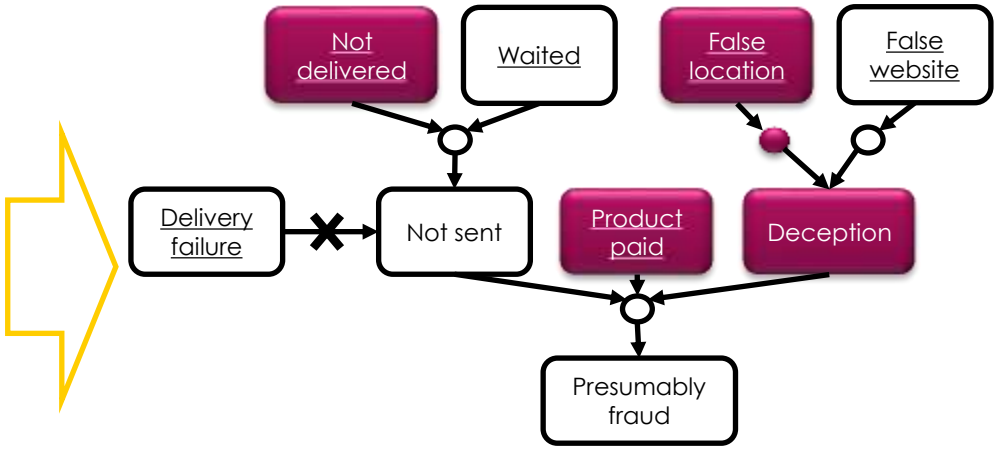
Fictitious example report 1
I would like to report fraud. I recently saw a bicycle for sale on eBay and contacted the advertiser. He said he lived far away, so he would send me the bike. I paid him in good faith, but have still not received anything. I saw on Facebook he lives nearby.

False location

Not delivered

Basic information extraction

Inferring possible fraud (or not)



Computational argumentation
Rules w. exceptions based on DCC & police policy rules

D. Odekerken, F. Bex, A. Borg, B. Testerink (2022) Approximating Stability for Applied Argument-based Inquiry. *Intelligent Systems with Applications*.

AI for intake – asking the right questions

Extracting observations from complaint form

Paid

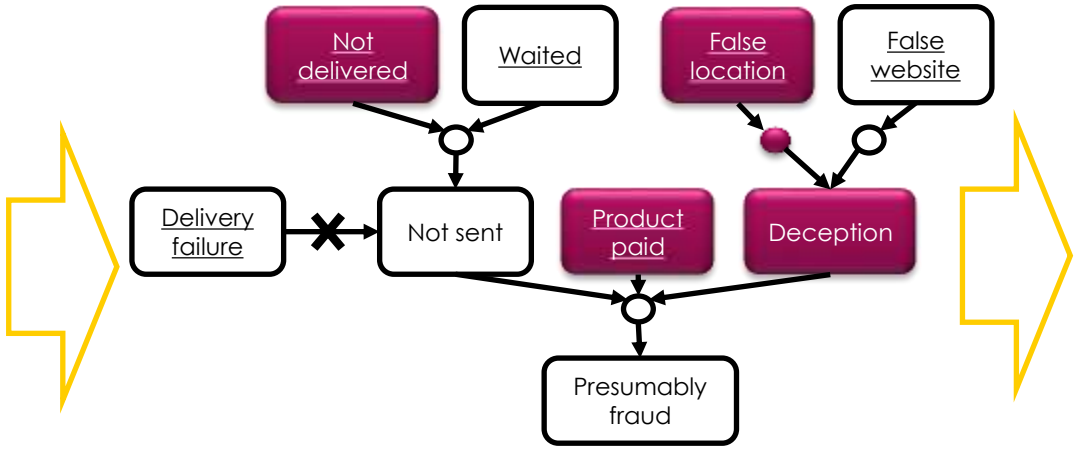
Fictitious example report 1
I would like to report fraud. I recently saw a bicycle for sale on eBay and contacted the advertiser. He said he lived far away, so he would send me the bike. I paid him in good faith, but have still not received anything. I saw on Facebook he lives nearby.

False location

Not delivered

Basic information extraction

Inferring possible fraud (or not)



Computational argumentation
Rules w. exceptions based on DCC & police policy rules

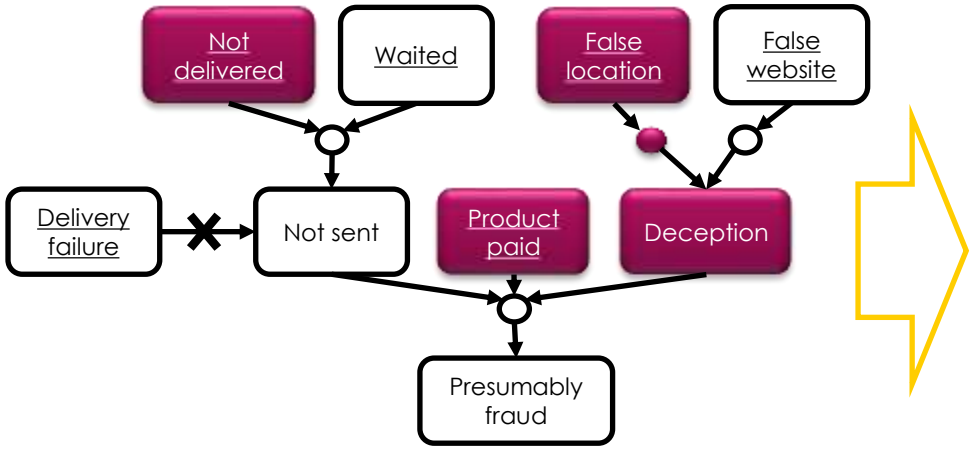
Asking for missing observations

Approximation algorithms
Can new info still change the conclusion (and if so which)?

A. Borg & F. Bex (2021) Explaining Arguments at the Dutch National Police. *Explainable AI for Law (XAILA)*.

AI for intake - explanations

Inferring possible fraud (or not)



Response
Thank you for your complaint. In your case, the system has concluded that it is not a case of fraud, since you did not wait for at least 5 days. We recommend you do not file an official report at this point.

Computational argumentation
Rules w. exceptions based on DCC & police policy rules

Explanations
Explaining (non-)acceptance in terms of arguments and counterarguments

AI for intake - evaluation

- Evaluate accuracy, user satisfaction
- Investigate citizen trust in automatic recommendations
  - How do users perceive recommendations by the system?
  - Do explanations matter?

E. Nieuwenhuizen, A. Meijer, F. Bex, S. Grimmelikhuijsen Explanations increase citizen trust in police algorithmic recommender systems: Findings from two experimental tests. *Under Review*

AI for intake – citizen trust & explanations

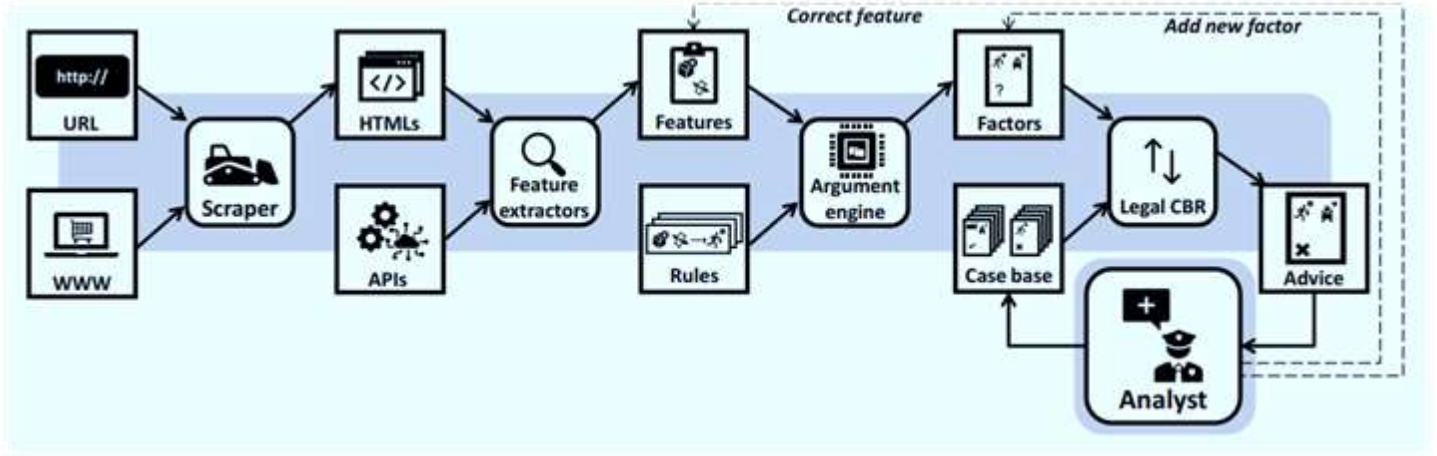
- Do citizens trust the system with and without an explanation?
  - Controlled experiments 1700+ participants
- Not fraud – still file an official report? (trusting behaviour)?
  - No explanation (control): 40-60% still filed report

E. Nieuwenhuizen, A. Meijer, F. Bex, S. Grimmelikhuijsen Explanations increase citizen trust in police algorithmic recommender systems: Findings from two experimental tests. *Under Review*

AI for intake – citizen trust & explanations

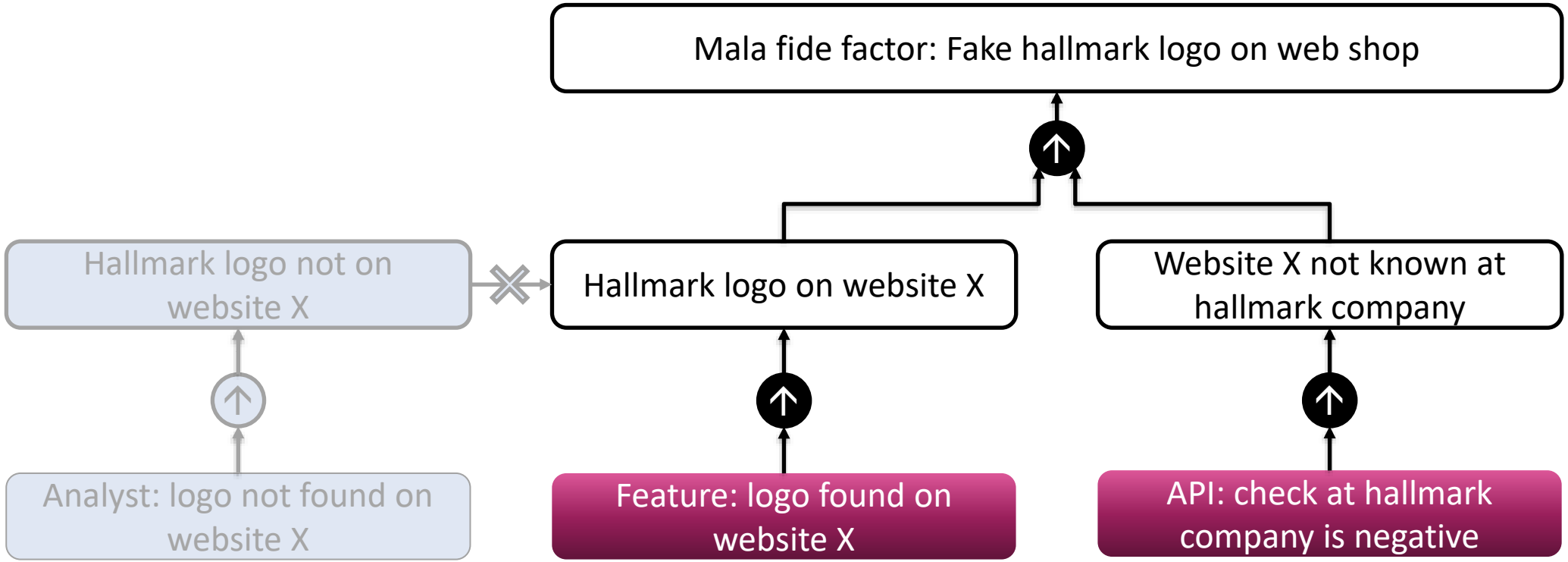
- Do citizens trust the system with and without an explanation?
  - Controlled experiments 1700+ participants
- Not fraud – still file an official report? (trusting behaviour)?
  - No explanation (control): 40-60% still filed report
  - With explanation: only 20-35% still filed report

Classifying web shops



- Webshop websites are scraped from internet
- Features (e.g. address, bank account, logo) are automatically identified by AI
  - Machine learning/data-driven AI
- Based on features it is determined if webshop has malafide (bad) and/or bonafide (good) factors
  - Knowledge-based argumentation

shopping secure



Scenarios about fraudulent web shops

- Cases or scenarios are of different types
  - Mala fide (bad) web shop
  - Bona fide (good) web shop
  - New cases are classified by comparing them to earlier cases

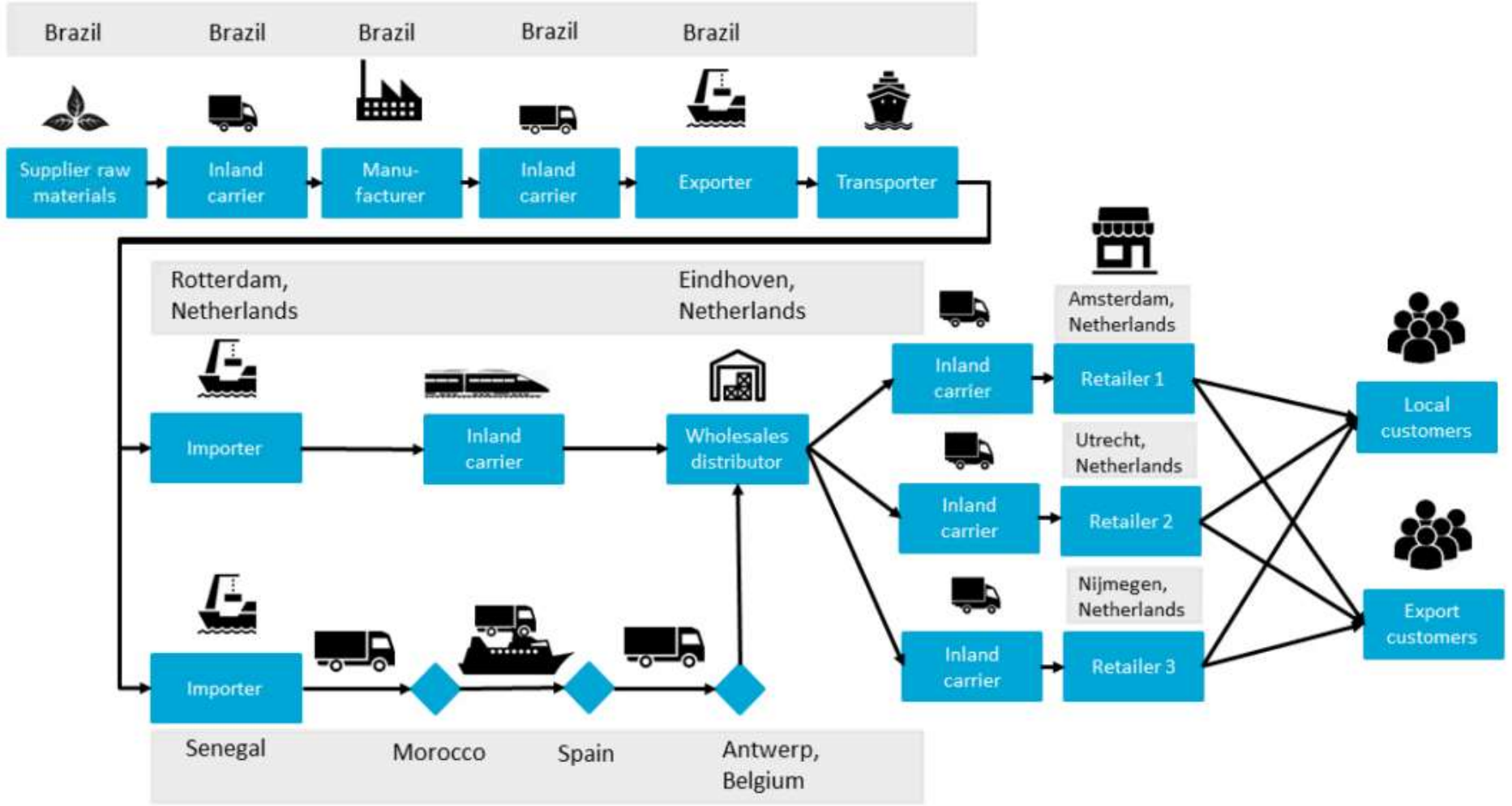
The police warns for the web shop www.mala-fide-web-shop.com:
1. They sell products for unrealistically low prices;
2. The Chamber of Commerce number does not exist;
3. The VAT number is invalid;
4. Registration date domain does not comply with date in terms & conditions.

Evaluation

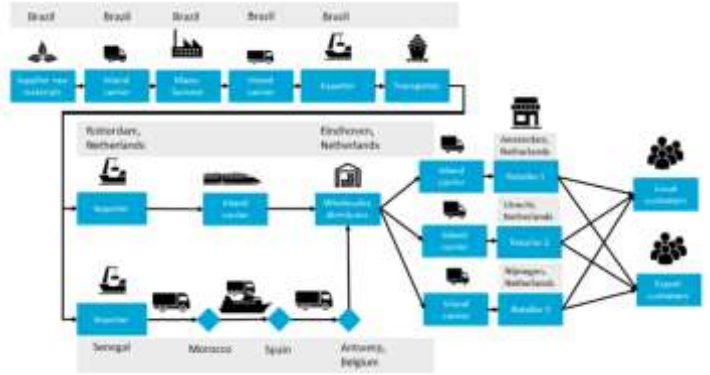
- Informal evaluation
  - 90% accurate classifications of mala-fide
- Useful, but work processes need to change
  - “I have to check the website myself anyway”

Knowledge-based AI for reasoning about criminal markets

Knowledge-based AI for criminal markets



Knowledge-based AI for criminal markets



- Knowledge about drug crime logistics & markets
- Can point to missing links, allow us to infer conclusions
  - E.g. we have a retailer and an importer, but are missing an inland distributor
  - E.g. If someone contacted both a wholesaler and a client, they are a retailer

Data-driven (machine learning) AI for search

Raw Data at the Police

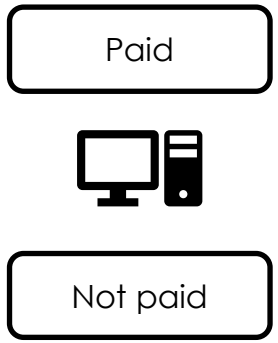
- Data comes from several sources
  - Citizens
    - Complaints, Witness testimonies, Open Public Data (e.g., social media, news, webshops)
  - Police officers & forensic scientists
    - Procès-verbal, Incident reports, Investigative reports, Lab reports, Internal communication / documentation
  - Suspects:
    - E.g., Data from Seized Data Carriers

Text classification

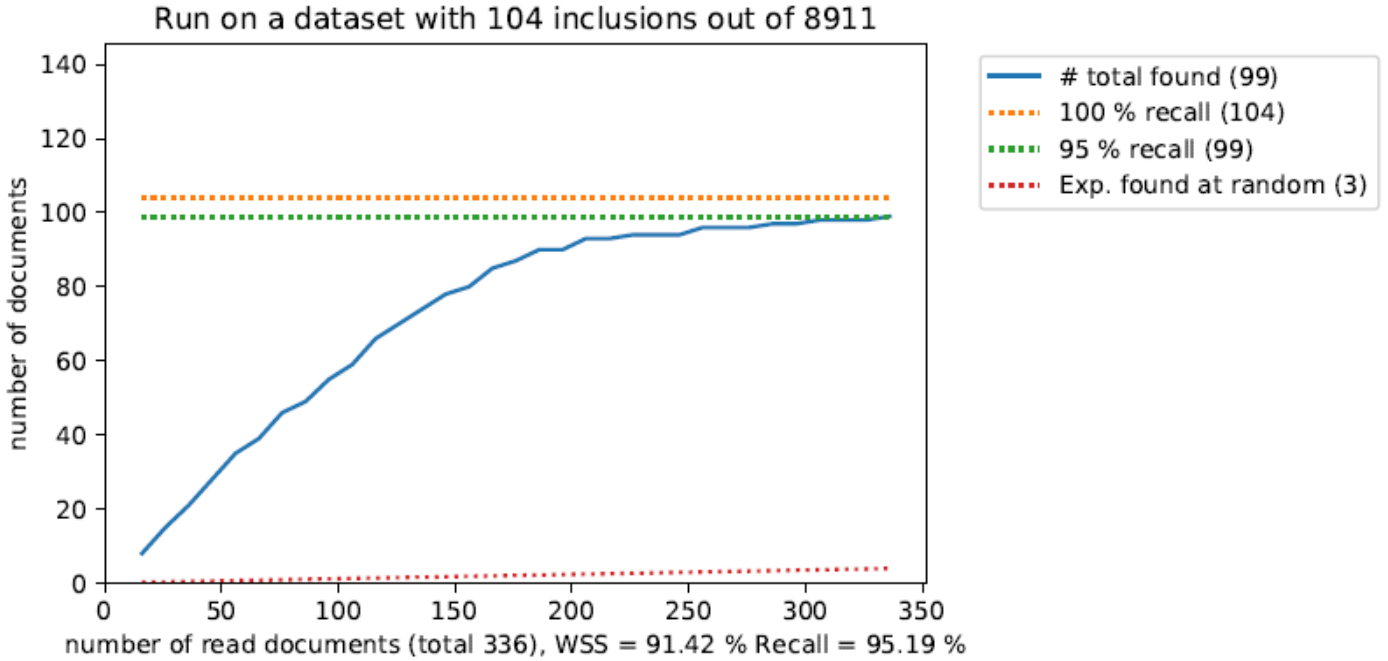
- AI can automatically classify documents
  - Paid – not paid
  - Threat – no threat
  - Relevant – not relevant

Fictitious example report 1
I would like to report fraud. I recently saw a bicycle for sale on Marktplaats and contacted the advertiser. He said he lived in Groningen, so he would send me the bike. I paid him in good faith, but have still not received anything. I saw on Facebook he lives in Maastricht.

Fictitious example report 2
I wanted to buy a bottle of Dom Perignon and a bottle of Crystal 1999 from John Doe via Marktplaats. Up to now, I have not received anything and John Doe does not respond to my e-mails.



Technology assisted review



- Search and find relevant documents in large dataset
  - E.g. criminal communication, case files
  - Whether it is relevant or not: automatic text classification
- *Active learning:* system proposes documents one by one and asks human if it is relevant.
  - Learns what kind of documents are relevant

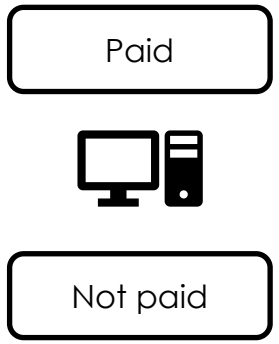
Explainable machine learning at the police

AI for explainable text classification

- Text classification for search & use in AI systems

Fictitious example report 1
I would like to report fraud. I recently saw a bicycle for sale on Marktplaats and contacted the advertiser. He said he lived in Groningen, so he would send me the bike. I paid him in good faith, but have still not received anything. I saw on Facebook he lives in Maastricht.

Fictitious example report 2
I wanted to buy a bottle of Dom Perignon and a bottle of Crystal 1999 from John Doe via Marktplaats. Up to now, I have not received anything and John Doe does not respond to my e-mails, so I haven't transferred the money yet.



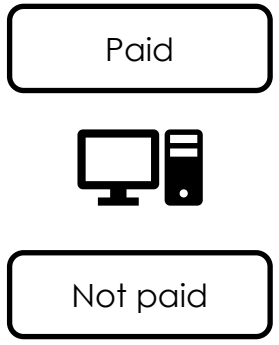
E. Herrewijnen, D. Nguyen, J. Mense & F. Bex (2021)
Machine-annotated Rationales: Faithfully Explaining
Text Classification. *Explainable Agency in AI Workshop*.

AI for explainable text classification

- Explaining text classification: Why did the AI classify the text as such?
  - Using *machine generated rationales* (highlighted sentences)

Fictitious example report 1
I would like to report fraud. I recently saw a bicycle for sale on Marktplaats and contacted the advertiser. He said he lived in Groningen, so he would send me the bike. I paid him in good faith, but have still not received anything. I saw on Facebook he lives in Maastricht.

Fictitious example report 2
I wanted to buy a bottle of Dom Perignon and a bottle of Crystal 1999 from John Doe via Marktplaats. Up to now, I have not received anything and John Doe does not respond to my e-mails, so I haven't transferred the money yet.

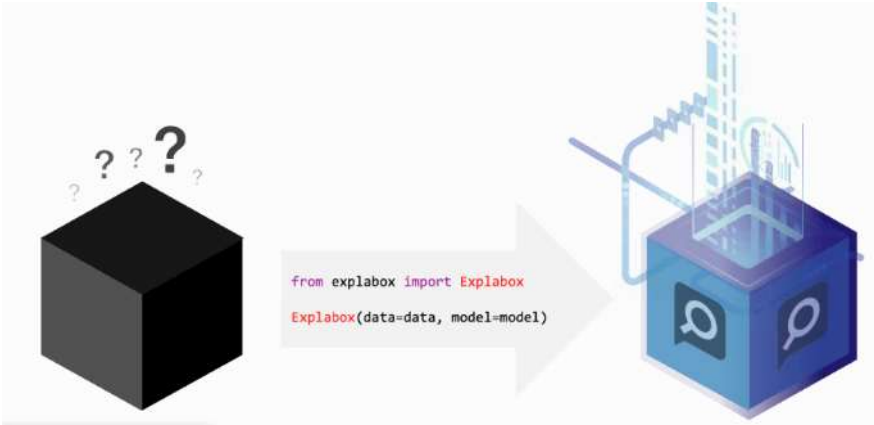


Fictitious example report 1
I would like to report fraud. I recently saw a bicycle for sale on Marktplaats and contacted the advertiser. He said he lived in Groningen, so he would send me the bike. I paid him in good faith, but have still not received anything. I saw on Facebook he lives in Maastricht.

Fictitious example report 2
I wanted to buy a bottle of Dom Perignon and a bottle of Crystal 1999 from John Doe via Marktplaats. Up to now, I have not received anything and John Doe does not respond to my e-mails, so I haven't transferred the money yet.

M. Robeer et al. (2023) The Explabox: Responsible AI Development & Usage. *Under review.*

Explainable AI for legal decisions



https://explabox.readthedocs.io/

- Open-source libraries & toolkit for AI model inspection
  - Data statistics
  - XAI: rationales, counterfactuals, LIME/SHAP
  - Robustness: spelling mistakes, typo's
  - Biases: names, gender, etc.
- A holistic view on the AI system
  - What kind of data? How (good) does the system perform? Why does the system do what it does?

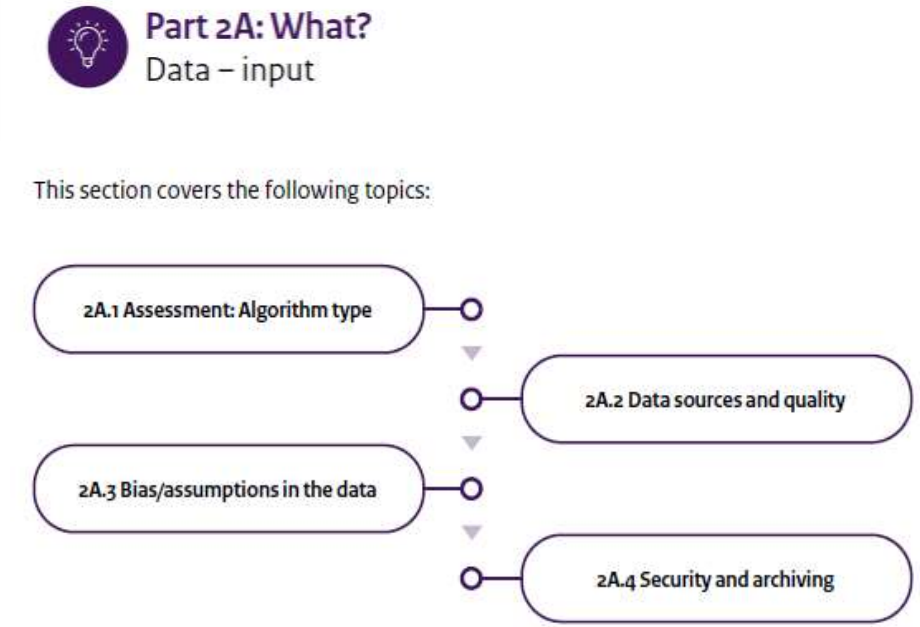
Explabox as assessment aid

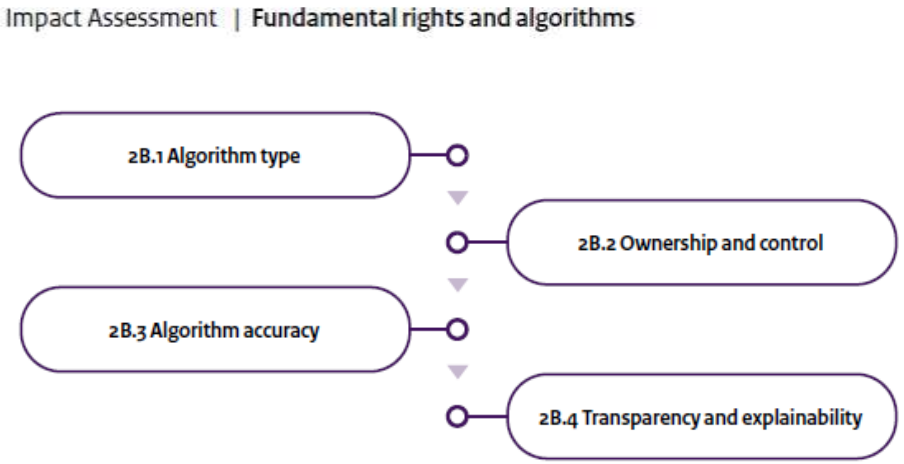
- Use information from Explabox for assessment
  - What kind of data? How (good) does the system perform? Why does the system do what it does?

Ministry of the Interior and Kingdom Relations

Impact Assessment
Fundamental rights and algorithms

Prototype stage at the police



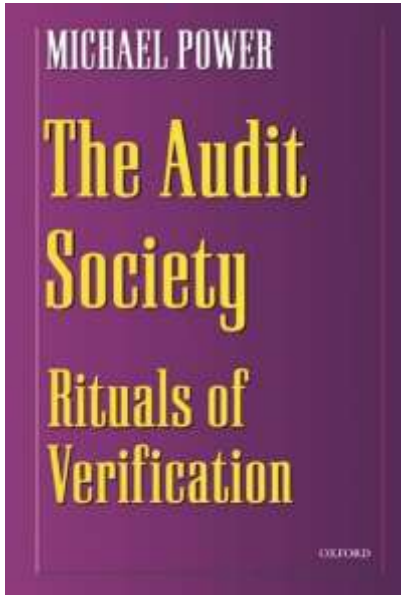


Ministry of the Interior and Kingdom Relations

Impact Assessment
Fundamental rights and algorithms

Rules, tools, and metrics

- Tools & metrics
  - What use are they? Intended and actual effects?
- New roles and responsibilities in organisations
- New research just started



Wieringa (2020) What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability. *FAT* 2020*.

algosoc Justice

Explainable AI for legal decisions


EUROPEAN COMMISSION

Brussels, 21.4.2021
COM(2021) 206 final

2021/0106 (COD)

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

- Rules: Operationalising transparency and contestability in the law
  - Equality of arms
  - Evaluating evidence and motivating decisions
- New research just started

Quattrocolo et al. (2020) Technical solutions for legal challenges: equality of arms in criminal proceedings. *Global Jurist*.

Bibal et al. (2021) Legal requirements on explainability in machine learning. *AI & Law Journal*

From Multimodal Data to Trustworthy Evidence in Court
AI4Intelligence

Explainable AI for legal decisions


EUROPEAN COMMISSION

Brussels, 21.4.2021
COM(2021) 206 final

2021/0106 (COD)

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

- Rules: Operationalising transparency and contestability in the law
  - Equality of arms
  - Evaluating evidence and motivating decisions

Bibal et al. (2021) Legal requirements on explainability in machine learning. *AI & Law Journal*

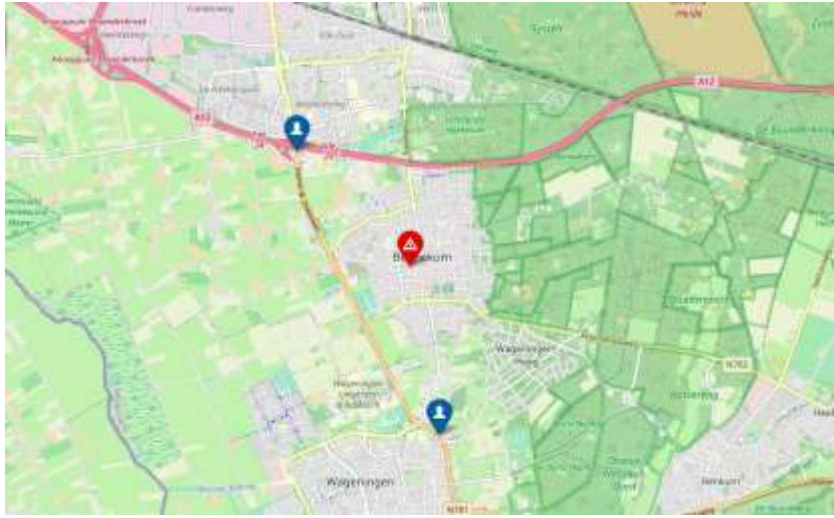
Almada (2019) Human intervention in automated decision-making: Toward the construction of contestable systems. *ICAIL 2019*.

Atkinson, Bench-Capon & Bollegala (2020) Explanation in AI and law: Past, present and future. *AI Journal*.

F.J. Bex (2011) *Arguments, Stories and Criminal Evidence: A Formal Hybrid Theory*.

Evaluating AI at the police

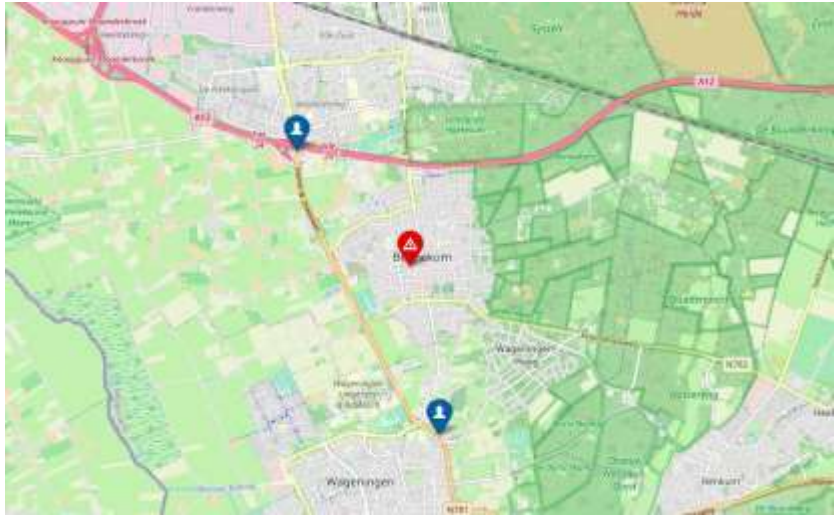
AI for police interception



- Notification of crime (e.g., robbery, smash & grab) and fleeing suspects
- Using knowledge about suspect behaviour, roads, etc., predict the suspect's route

Van Droffelaar, I.S., Kwakkel, J.H., Mense, J.P., Verbraeck, A. (2022)
Simulation-optimization configurations for fugitive interception.
Proceedings of the 2022 Winter Simulation Conference.

Example 3 - AI for police interception



- Notification of crime (e.g. robbery, smash & grab) and fleeing suspects
- Using knowledge about suspect behaviour, roads, etc., predict the suspect's route

- "Just like I thought"
  - Expert dispatchers only followed the recommendations of the system if they coincided with their own intuitions
  - Explanations hardly influence whether they trust/follow the recommendation

Van Droffelaar, I.S., Kwakkel, J.H., Mense, J.P., Verbraeck, A. (2022) Simulation-optimization configurations for fugitive interception. *Proceedings of the 2022 Winter Simulation Conference*.

F. Selten, M. Robeer, S. Grimmelikhuijsen (2022) 'Just like I thought'. Street-level bureaucrats trust AI recommendations if they confirm their professional judgement. *Public Administration Review*.

AI at the police – concluding thoughts

Concluding

- AI & Law in practice – while doing research
  - Involve practitioners
  - Evaluate broadly with different disciplines
- Machine learning is not the answer to everything!
  - Good for “sensing” in noisy data (free text, images, websites)
- Argumentation/reasoning for drawing conclusions
  - Transparent, based on law & policy, insights into criminal behaviour