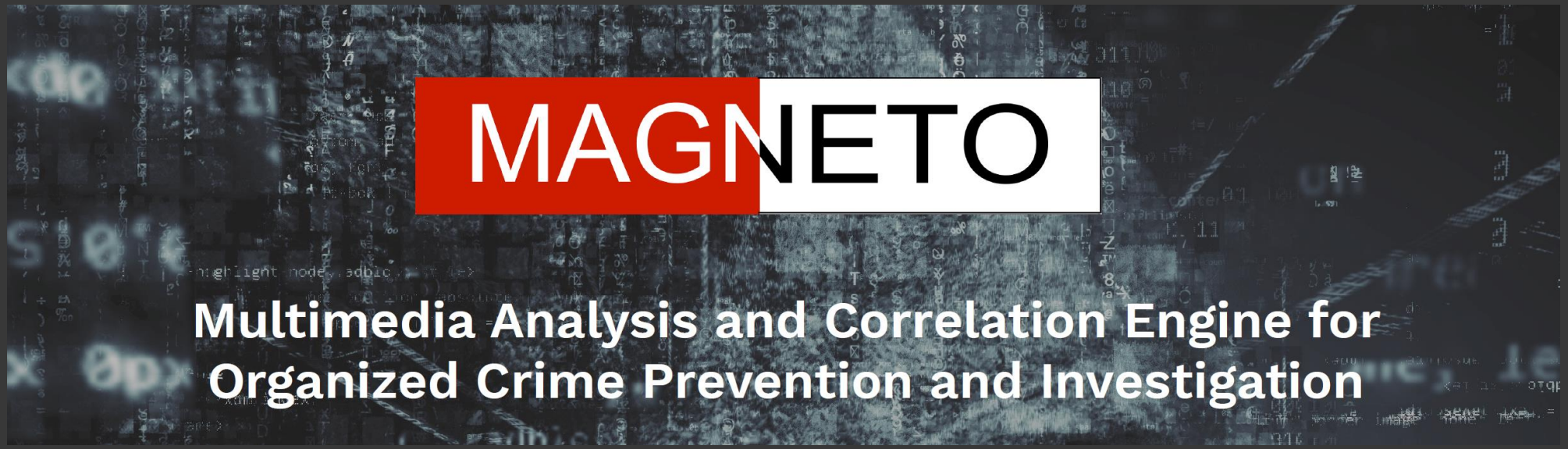
MAGNETO



Artificial Intelligence for Law Enforcement: Legal & Ethical Considerations & Strategies in MAGNETO

Thomas Marquenie, KU Leuven – Centre for IT & IP law
thomas.marquenie@kuleuven.be

HORIZON 2020

MAGNETO

Introduction

MAGNETO

Introduction

AI applications in law enforcement

- **Big data analytics**
  - Establish (hidden) trends and correlations in large volumes of data
- **Predictive policing**
  - Identify locations, persons and circumstances likely to be involved in crime
- **Robotics, drones and cameras**
  - Real-time facial recognition, gait analysis, augmented reality
- **Algorithmic profiling**
  - Risk assessments and behavioral analytics

Augment human decision-making & Inform police strategy

MAGNETO

Consequences of Police AI

- Efficient, fast and accurate analysis of large datasets
- Effective allocation of resources and manpower
- Supporting police personnel by enhancing cognitive abilities
- Applied risk assessments and data-driven strategies

- Risks to human rights and individual liberties
- Cementation of bias and discriminatory practices
- Chilling effects on citizen behavior
- Possibility of function creep with novel technologies

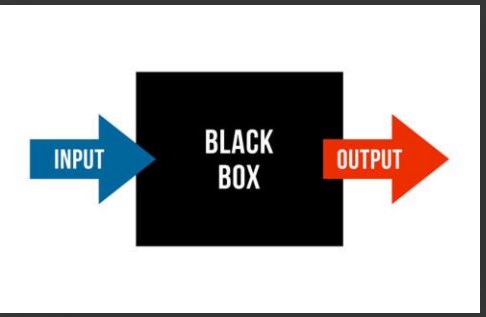
MAGNETO

Legal & Ethical Considerations

MAGNETO

Algorithmic Fairness & Accountability

- **Discrimination & Bias**
  - Algorithms are "taught" to interpret data based on our perceptions
  - Replication and institutionalization of prejudicial practices
- **Technical inaccuracies**
  - "Garbage in, garbage out"
  - Faulty designs or model shortcomings
- **Opacity & Unaccountability**
  - "Black box" systems are inscrutable to both users and citizens
  - Lack of transparency and inability to provide satisfactory justifications



<u>Unintended & harmful results that are difficult to detect and challenge</u>

HORIZON 2020

MAGNETO

Data Protection

- **European Union – Law Enforcement Directive 2016/680**
  - Domestic activities now subject to European DP rules (in effect: 2018)

- **Regulates processing of "personal data" in policing / criminal justice**
  - Information regarding an identified / identifiable natural person
- **Sets out general principles for data collection, storage, analysis and transfer**
  - Purpose limitation, data minimization, fairness / lawfulness, security
- **Places certain responsibilities and behavioral standards on LEAs**
  - Profiling, impact assessments, sensitive data, disclose information…
- **Awards citizens with right to (some) control over their data**

MAGNETO

Human Rights

- **Privacy** (art. 7-8 EU Charter, 8 ECHR)
  - Protection against unreasonable interferences in private life and personal data
  - Limits on police data use (contrast with big data analytics)
- **Equal treatment** (art. 21 Charter, 14 ECHR)
  - All persons are treated equally before the law, without discrimination
- **Fair Trial** (art. 47 Charter, 6 ECHR)
  - Presumption of innocence
  - Equality of arms (access to evidence & reasoning behind the charges)
  - Overall fairness of the trial (including other rights)
- **Limitations**
  - Proportionality, objective / reasonable justification, public interest, safeguards

Tensions with novel police technology

HORIZON 2020

MAGNETO

MAGNETO Mitigation Strategies

MAGNETO

MAGNETO Approach

Relevant strategies for all legal principles

Extensive legal and ethical **guidelines** (and **translate** them **into design strategies**)

Active involvement of and **interactions between** different **stakeholders** (such as system developers, end-users, legal and ethics specialists)

Extensive **training and awareness raising** activities for end-users

MAGNETO

Initial Design

- **Review of the legal & ethical framework**
  - Drafting of general guidelines in collaboration with LEA's / technical partners
- **Consider legal-ethical risks at the design and development phase**
  - Human rights, data protection, ethical / societal values
  - Explore safeguards and counter-measures at the beginning
- **Data Protection Impact Assessment**
  - Continuous process
  - Interim & Final DPIA
- **Start with the most secure and privacy-oriented option by default**

MAGNETO

Processing Principles

- **Purpose Limitation**
  - Database integration / restrictions to limit processing to specified & explicit purposes
- **Data Minimization**
  - Access controls for use of relevant, adequate and non-excessive data processing
  - Limiting certain functionalities / datasets to specific user profiles
  - Pseudonymisation/anonymization of less/non relevant data
- **Accuracy**
  - Alerts to users to preserve accuracy / update information
  - Easily rectifiable/updateable information
- **Security (integrity and confidentiality)**
  - Access, authorization and user control, communication control, input control
  - Recovery and back-up

HORIZON 2020

MAGNETO

Data Protection Standards

- **Time limits for storage & review**
  - Built-in processes with reminders to reflect legal / organizational periods
- **Distinction between categories of data & data subject**
  - Automated / manual tagging of different types of data to enable distinction
- **Sensitive categories of data**
  - Restricted processing operations + tagging of data
- **Record-keeping and logging**
  - Of all system processes and user interactions – enabled by the tools themselves
- **“Automated decision-making”**
  - Prohibition on certain categories of data, human in the loop, explanation

HORIZON 2020

MAGNETO

Accountability & Fairness

- High quality, diverse, large **training datasets**
- System **testing and validation** to confirm accuracy and lack of bias
- **Explainable and understandable system processes**
  - Enabling assessment of the reasoning behind specific actions, decisions & outcomes
- **Foreseeability and verifiability** of system behaviors and actions
- **Auditable documentation** on algorithms, datasets, machine learning methods
- **Emphasize human oversight**
  - Validate outcomes of decision-making (meaningful human intervention)
- **"Whitelisting" and "blacklisting" rules** that the system should always follow or never disregard

MAGNETO

Concluding Remarks

MAGNETO

Conclusion

1. **Conducting impact assessments and mitigation measures**
2. **Value-sensitive design and deployment**
   - Counter bias in datasets (data quality) & validating methodological fairness
3. **Prioritizing transparency and accountability** (+ redress)
4. **Following data protection principles**
5. **Emphasizing a capable human in the loop**
   - Avoid rubberstamping and decisions made solely on the basis of algorithms
   - Ensure direct human control over “reasonable suspicion” and outcomes
6. **Establishing independent oversight and auditing**
7. **Adhering to established human rights standards**
   - Objectivity, proportionality, pertinence

HORIZON 2020

MAGNETO

Thank you

Acknowledgment

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 786629.

Get in touch: @H2020Magneto MAGNETO H2020 http://www.magneto-h2020.eu

Coordinator contact: Dr. Konstantinos Demestichas - cdemest@cn.ntua.gr