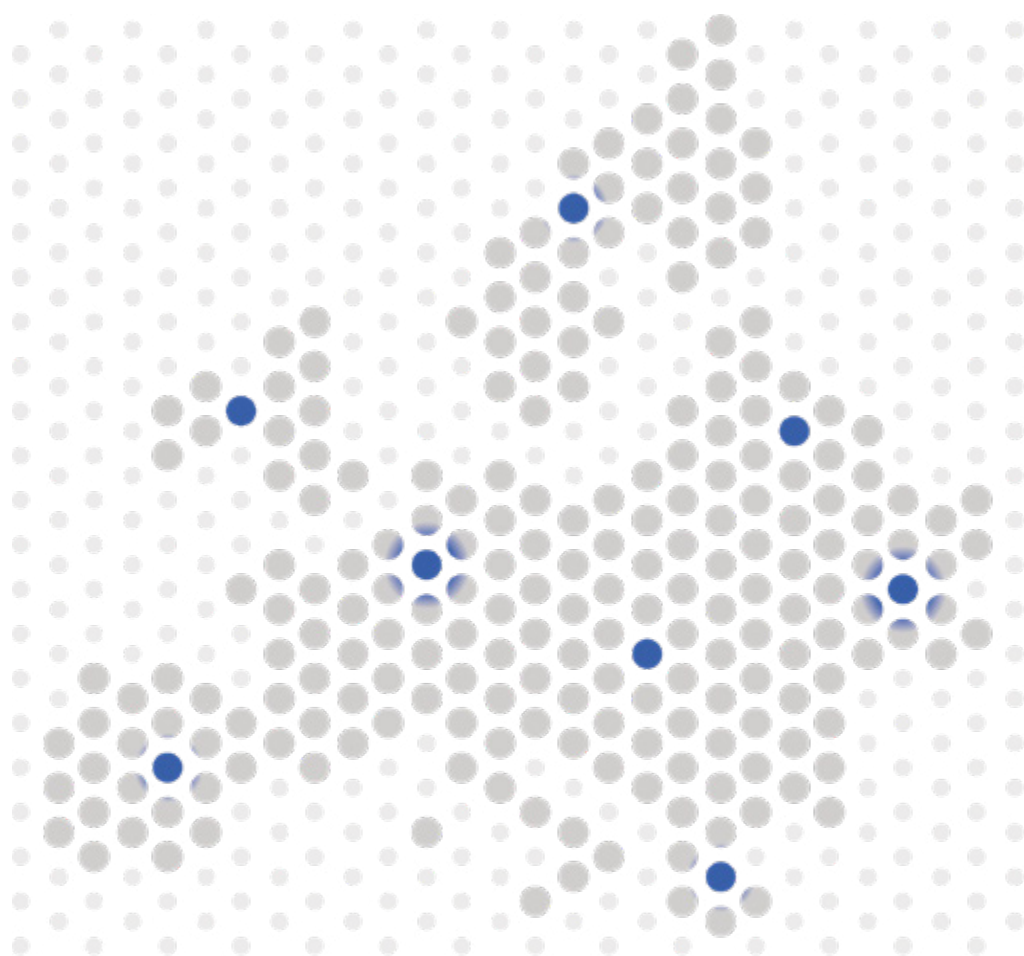


NWSLTR

TOWARDS AN EVIDENCE-BASED MODEL FOR BIG DATA POLICING



BIGDATPOL

OPINION

The importance
of data (quality)

3

HIGHLIGHTED

PUBLICATION
Mobile phone
data

4

RESULTS

Experimental
study police
zone Zennevalei

5

WHAT ABOUT THE AI ACT?

7

BUILDING THE NETWORK

8

EDITORIAL PROF. DR. WIM HARDYNS

Welcome to the second edition of our BIGDATPOL newsletter!

In this newsletter, we are excited to share some updates on the progress of our Big Data Policing research programme.

Our project is moving forward and I am proud to present some of the initial results.

The pilot project in the police zone of Zennevallei in Belgium has provided valuable insights into the use of predictive models. The dedication and willingness of the police staff to engage with these new tools is admirable. However, there are still challenges to address, particularly in improving the integration of data into routine police operations.

Over the coming months, we will continue to refine and extend the initial model, carefully considering potential pitfalls from both a practical and ethical perspective. Our aim is to ensure compliance with the AI Act and other relevant regulations, while minimising bias through rigorous data selection and quality assessment.

Last but not least, I am delighted to announce the launch of our new website! Visit us online to explore our research projects, latest updates and opportunities for collaboration:
www.bigdatpol.com.

Please feel free to share this newsletter with your network and contact us if you would like further information or would like to explore potential collaborations.

Kind regards,

PROF DR WIM HARDYNS
PRINCIPAL INVESTIGATOR BIGDATPOL



OPINION



THE IMPORTANCE OF DATA (QUALITY) IN WORKING WITH AI

By Dr. THOM SNAPHAAN,
postdoctoral researcher
BIGDATPOL

After the steam engine, electricity and the internal combustion engine, artificial intelligence (AI) is now considered as a systems technology¹. In other words, it is not just a technological application, but technology that is used by society and the economy as a whole for a variety of purposes. In this opinion piece, I will support and explain the importance of data in AI applications.

Blame it on the data

It is not uncommon for AI and algorithms to be referred to in negative terms. They are usually conjugated with adjectives such as 'discriminatory' or, more generally, 'biased'. However, it is a misconception to think that the problems that these algorithms sometimes cause are due to the algorithms themselves or to the technology in general. Especially with self-learning algorithms, the problem is not so much with the algorithm, but with the data used to train it. So it is a misconception to blame the algorithm.

Garbage in = garbage out

Building on this, the well-known adage garbage in = garbage out applies. If biased data are used to train the algorithm, the results (and the algorithm itself) will be biased. So the solution is not to change the algorithm itself, or even to ignore the results. If an algorithm produces undesirable results, this is also a signal for change. It reveals problems that were already present in the data (and the practices they reflect), such as discriminatory or, more generally, biased behaviour. It is a signal to decide: we need to do this differently.

Data do not speak for themselves

A final misconception I would like to dispel is that data, aided in part by AI, can speak for themselves - like a chef watching a piece of meat cook itself². No one will ever claim that a piece of meat cooks itself, or that a particular piece of meat implicitly leads to a particular dish. There is always a chef who decides which piece of meat to buy, how to prepare it, and what the final result should be in terms of composition and aesthetics. The opposite seems to be true of data; data talks and we listen. But nothing could be further from the truth. What is true of meat is also true of data: there is always someone 'turning the knobs', making decisions throughout the process. For example, choices are made in the selection of data sources or in the selection (or construction) of variables that determine the final results³.

In conclusion, I believe that more attention should be paid to data quality when working with AI in policing and security practice. Frameworks are available to provide practical guidance on how to assess data quality⁴. Finally, I would like to note that technology is simply not always the solution, nor does it have to be. AI seems to be the modern hammer that 'sees' a nail in everything and at the same time 'thinks' that everything needs a nail.

REFERENCES

1: Wetenschappelijke Raad voor Regeringsbeleid. (2021, 11 november). *Opgave AI. De nieuwe systeemtechnologie*. <https://www.wrr.nl/publicaties/rapporten/2021/11/11/opgave-ai-de-nieuwe-systeemtechnologie>

2: See Balazka, D., & Rodighiero, D. (2020). Big data and the little big bang: An epistemological (r)evolution. *Frontiers in Big Data*, 3, Article 31. <http://doi.org/10.3389/fdata.2020.00031>

3: Snaphaan, T., Hardyns, W., & Pauwels, L. J. R. (2023). Hypotheses non fingo? De rol van theoretische verklaringen in een bigdatatijdperk. In T. Snaphaan, W. Hardyns, A. J. van Dijk, R. Spithoven & R. Van Brakel (Eds.), *Big data policing* (pp. 33-52). Gompel&Svacina.

4: Snaphaan, T., Hardyns, W., & Pauwels, L. (2024). Expanding the methodological toolkit of criminology and criminal justice with the Total Error Framework. *Journal of Crime & Justice*, 47(1), 112-129. <https://doi.org/10.1080/0735648X.2022.2114099>

This article has been published in Dutch and French as respectively Snaphaan, T. (2024). Het belang van (kwaliteitsvolle) data bij het werken met AI. *Blue Minds*, 5(3), 36-37; Snaphaan, T. (2024). L'importance des données (de qualité) lorsqu'on travaille avec l'IA. *Blue Minds*, 5(3), 36-37.

HIGHLIGHTED PUBLICATION

'Do mobile phone data provide a better denominator in crime rates and improve spatiotemporal predictions of crime?'

In 2021, we published a scientific article showing the added value of mobile phone data in calculating crime ratios and for predicting time-spatial concentrations of crime. This article appeared in the ISPRS International Journal of Geo-Information and was written by Anneleen Rummens, Thom Snaphaan, Nico Van de Weghe, Dirk Van den Poel, Lieven Pauwels and Wim Hardyns.

This study investigated whether the use of mobile phone data provides a better picture when calculating crime ratios and whether it can improve time-spatial predictions of crime compared to (commonly used) data on the residential population. In this study, we used mobile phone data to measure the so-called 'ambient population'. This ambient population refers to the number of people present in a given area at a given time. The hypothesis is that these data provide a more dynamic picture than residential population data (i.e., the number of people living in a given area; which is often made available annually), and are therefore useful in analysing crime types with a mobile target.

First, the extent to which crime ratios differ based on the two population measures used was examined. A crime ratio is a statistic often used to express crime in a relative sense, as it helps reveal crime concentrations over time and space based on an underlying risk population. For crimes with a mobile target - which we included in this study - such as aggressive theft, battery, and bicycle theft, the ambient population was found to identify totally different hotspots than those calculated using the residential population. This is because the patterns of the ambient population and the residential population are very different (see Figure 1). In general, stronger correlations were found between offences and the ambient population, than with offences and residential population.

Second, it was examined whether the use of the ambient population also led to better predictions of crime over time and space. Here, it was found that the models with the ambient population performed structurally better than the models with the residential population.

Here are three key conclusions from this study:

- 1. Stronger correlation with crime:** The ambient population, based on mobile phone data, shows a stronger correlation with recorded crime of aggressive theft, battery, and bicycle theft than the traditional residential population.
- 2. Different crime patterns:** When crime rates are calculated based on the ambient population, different hotspots are identified compared to crime rates based on the residential population. This highlights the dynamics of crime at different times and locations, thus underlining the importance of taking this into account in crime analyses.
- 3. Improved prediction models:** The use of mobile phone data in prediction models for time-spatial concentrations of crime significantly improves the accuracy of these models. This provides police and policymakers with a more powerful tool to predict crime and take preventive action.

The full article is available [online](#).

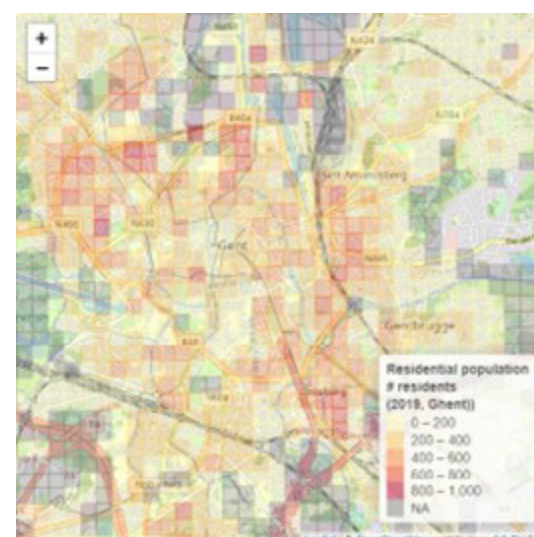
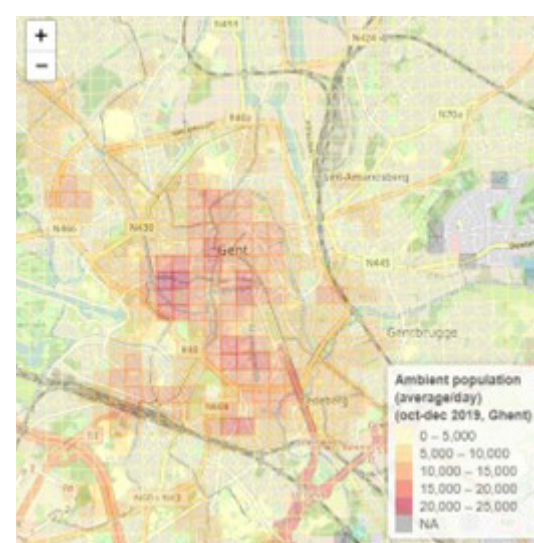


Figure 1: Different hotspots of residential population (top) and ambient population (bottom). Image obtained from Snaphaan et al. (2024).



PUBLICATIONS

Interested in more? Check out our reading list on biblio and discover the research that has been published by the team of Prof Dr Wim Hardyns.



An exploratory experimental study on the implementation of big data policing within the Flemish police zone Zennevallei

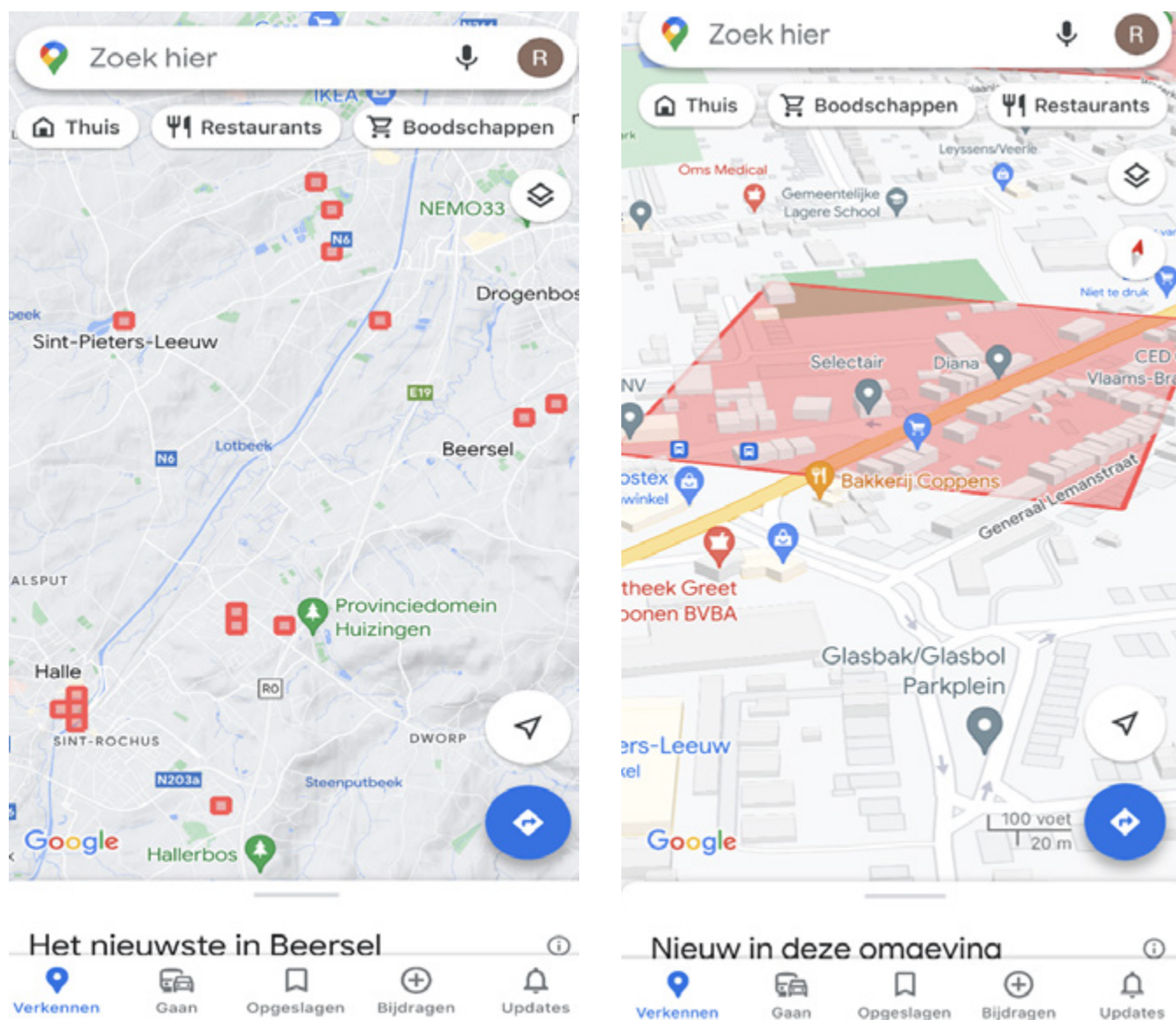


Figure 1: Example of a hotspot map used during the experiment.

In 2023, the BIGDATPOL team conducted an exploratory experimental study on the implementation of big data policing within the Police Zone (PZ) Zennevallei. Over a six-month period, a predictive crime model was used to predict where and when the likelihood of future residential burglaries was high, with the aim of making police patrols more targeted and proactive in the identified high-risk areas.

More specifically, the study aimed to evaluate both the effectiveness and practicality of big data policing within police work, taking into account the current use of police patrols and the contemporary context of policing.

For this purpose, an experimental research design was applied, dividing the police zone into experimental areas and control areas. On a weekly basis, the predictive crime model identified/predicted a number of high-risk areas where future residential burglary was most likely to occur, for both the experimental and control areas.

However, only the predictions for the experimental areas were made available to the police zone using digital hotspot maps (see Figure 1), so that officers in these areas could tailor their patrols to the predicted hotspots. The predictions for the control areas were not shared and served

only as a basis for comparison to assess the effectiveness of the intervention. This allowed a clear distinction to be made between outcomes in areas where patrols were matched based on the predictions and areas where usual patrols continued without predictions.

Based on this experimental research design, some quantitative parameters were then analysed, including the number of residential burglaries before, during and after the study period, as well as the level of police presence in the predicted risk areas. In addition, several qualitative aspects were also examined, including police officers' perceptions and

attitudes towards big data policing, their trust in the technology, the perceived impact on their work(load) and any ethical considerations.

Overall, the results of the study show that big data policing can indeed contribute to a more efficient allocation of police patrols, but that a lot of preconditions need to be met for big data policing to also effectively contribute to

- 1) deterring potential burglars and
- 2) measuring and demonstrating crime-reducing effects

Although the predictive model was found to be sufficiently accurate in identifying and predicting high-risk areas, some important factors were found to affect the effectiveness of the model.

For example, a thorough quantitative analysis of GPS data from police patrols showed that the level of police presence in the predicted high-risk areas within the experimental areas was not significantly different from the level of presence in the control areas.

This implies that the predictions alone were not sufficient to bring about a targeted increase in visible police presence in the risk areas. This significantly limited the effectiveness of the predictive model, as a more targeted deployment of police presence—a crucial condition for achieving deterrence and reduction effects—could not be realised in practice.

Based on focus groups, surveys and observations, several other important factors were identified that influenced the implementation of big data policing within the police zone. These factors included the level of acceptance and use of the predictions by police officers, the effectiveness of communication on the deployment of risk areas, and organisational challenges in integrating the predictive models into daily patrol practices.

Although police officers saw big data policing as supporting their work, depending on their role, they observed some challenges.

This was because the implementation of big data policing brought additional workload and restricted freedom of action while patrolling. Repeatedly patrolling the same high-risk areas led to feelings of boredom among police officers over time.

Moreover, due to the high workload, police officers did not always have the opportunity to consistently patrol the predicted risk areas. In particular, police officers working within the intervention service experienced significant difficulties in using the technology.

Future evaluation studies on big data policing should overcome the limitations of this evaluation study as much as possible. In that respect, it is clear that future studies should pay much more attention to how the predictions are made available to police districts and how police presence is deployed based on the predictions. Police officers in the field clearly need more guidance and feedback when time-spatial crime predictions are used to direct police patrols. In addition, it remains to be seen whether big data policing is a useful tool in the context of the regular operation of intervention services.

Indeed, intervention services are often faced with unforeseen and urgent situations that require immediate response, which may conflict with the objectives pursued under big data policing. Developing more flexible and proactive patrol strategies that better reflect the dynamics of high-risk areas can contribute to more effective deployment of technology.

Strengthening police officers' involvement in the design and implementation of such technologies can also help increase their confidence and improve the overall effectiveness of big data policing.



Police zone Zennevallei is considered a medium-sized police zone in Flanders accounting for almost 105 000 inhabitants.

BRUSSELS

KDCS WORKSHOP



With the AI Act having entered into force on August 1st, 2024, we are constantly reflecting how our actions are addressing the arising legal responsibility, and beyond, ensure our model is trustworthy and ethical.

We were consequently delighted that our PhD researcher Naomi Theinert could join the Policy Prototyping Workshop on 4 September, organised by the Kenniscentrum Data & Maatschappij (KDCS) and hosted by FARI – AI for the Common Good Institute in Brussels to represent our BIGDATPOL project.

Through collaborative team discussions and brainstorming tasks, Article 14 on Human Oversight of the AI Act was at the forefront of this session. By contextualising the requirements of Article 14 within our own project, this participatory workshop enabled us to design an initial prototype to help our project operationalise human oversight. In order to design a prototype for human oversight, we identified the end-users of BIGDATPOL, and mapped the potential fears, possibilities and thoughts that the users might have. Building on this mapping exercise, we pinpointed the risks associated with the system, while exploring solutions to address them.

In the upcoming month, we will continue finetuning the initial prototype by focusing on how to communicate and explain the model, while providing insights via infographics and progressing on how to train the users to read and understand the model output. To accommodate the risk of pitfalls in the model and unexpected incidents, incorporating a complaints mechanism is further crucial to allow for adjustments along the way.

WHAT ABOUT THE AI ACT?



The AI Act is the world's first comprehensive AI regulation. It groups AI applications by risk, with stricter rules for higher-risk applications.

The AI Act lays out specific requirements for the development and use of high-risk systems, such as:

- Establishment of a risk management system
- Data governance and management
- Technical documentation
- Automatic record-keeping
- Transparency and information provision to users
- Human oversight
- Ensuring accuracy, robustness, and cybersecurity. High-risk AI systems must undergo a third-party assessment before they can be sold or used (referred to as conformity assessment)

According to the AI act, AI-driven predictions of crime in time and space should be considered high-risk applications. What does this mean?

How do we consider these requirements in BIGDATPOL?

Rigorous research and adhering to academic standards, links naturally to the requirements under the AI Act. Specifically, since we incorporate a multi-dimensional approach and our key focus is on evaluation, the requirements align with our research goals.

Moreover, our research output will help in increasing the reach of both our methods and results. Feedback and reviews by stakeholders, as well as academics, will help us improve and yield robust results.

How are the requirements relevant for users?

End-users and stakeholders of our model will have the right to receive information that provides further insight onto what influenced the output of our model. What changes in the features would have yielded a different outcome? What features play a role? What does human oversight constitute for each decision-maker? What does human autonomy and oversight mean within BIGDATPOL?

Investigating these questions empirically and identifying explainability in practice, is one example of how the AI Act requirements will be addressed in our work.

BUT - the AI ACT is not a standalone legislation

The AI Act ties into and should be considered with existing legislation, such as the **GDPR** (General Data Protection Regulation), the **LED** (Law Enforcement Directive) and **Fundamental Rights Charter**.

In addition to the AI Act, our goal is to contextualise our actions in the existing legislative frameworks and ethical principles, mainly human autonomy, prevention of harm, explicability, and fairness.

Want to know more? Read the European Parliament's documentation on the EU AI act [here](#).

Do you have specific questions, concerns or points for improvement? Let our team know at bigdatpol@ugent.be.

BUILDING THE NETWORK

BUCHAREST ESC Conference Eurocrim 2024



From 11 to 14 September, the BIGDATPOL team was present at the Annual Conference of the European Society of Criminology in Bucharest.

We had the opportunity to present our research in pre-arranged panels:

1. Towards an evidence-based model for big data policing (BIGDATPOL)

BIGDATPOL aims to harmonise statistical, criminological, economic, legal and ethical facets into an evidence-based model, that contributes to a safer and fairer future for our communities. This approach uses historical data to predict and pre-empt potential crime hotspots, optimising police resources and ultimately reducing crime rates. In this thematic panel, we presented findings from several studies related to the overarching programme, each examined from a different disciplinary perspective.

- Professor Hardyns introduced the BIGDATPOL research programme to the full room. He gave a general overview and demonstrated the state-of-the-art in both research and practice regarding big data policing. He elaborated on the aims of the program and the way that the overarching program is divided into several tracks. Inge Claessens shared how we aim to bring together relevant stakeholders in this field in the future.
- 'Evaluating big data policing: The development of a comprehensive

quantitative and qualitative evaluation framework' by Thom Snaphaan. The presented framework integrates both quantitative and qualitative approaches to assess big data policing, focusing on ethics, community dynamics, criminological impacts, and cost-benefit analysis. It offers valuable insights for policymakers, law enforcement, and researchers to evaluate data-driven strategies in policing.

- Charlotte Vandenbrande addressed the 'Attitudes of police officers towards big data policing'. She presented the results of our surveys and focus groups in 20 Flemish police zones, exploring police officers' perceptions, expectations and experiences of big data policing. Key areas included effectiveness, trust, ethics and practical implementation, providing insights to inform the development of data-driven policing models.
- 'Exploring explainability in big data policing: Identifying feature importance in decision-making models' by Naomi Theinert. In her presentation she focused on identifying feature importance in decision-making models used in place-based policing. By examining explainability techniques, the study highlights the trade-offs between methods and emphasises the need for transparency and accountability in data-driven law enforcement practices.

- 'A temporal analysis of crime concentrations in micro-places and the ambient population' by Robin Khalfa. He presented research on temporal crime concentration, exploring the application of the Crime Heartbeat methodology (Rafael Prieto-Curiel) to examine when crime concentrates over a weekly cycle. The study examines the temporal concentration of different types of crime in Ghent, Belgium, between 2007 and 2018.

2. After the BIGDATPOL panel, Robin Khalfa and professor Hardyns presented their research on **car theft prediction and prevention** in a separate panel on car theft, discussing how big data and big data analytics can support car theft prevention at micro places.

This presentation examined the potential of a place-based big data policing approach for predicting and preventing car theft incidents. First, an overview was provided of the ways in which big data sources, including location technology data, municipal service records (e.g. parking registrations), and emerging sources like mobile phone data, can be used to quantify potential theft targets (i.e. vehicles) and their associated attributes. Moreover, the integration of predictive big data analytics for the prediction of car theft occurrences was discussed. The preliminary findings from a comparative analysis of diverse machine learning models applied to a car theft dataset spanning the period 2007-2018 for the city of Ghent, Belgium, were presented. Finally, the implications of the findings were discussed in light of different crime prevention strategies within the context of car theft.

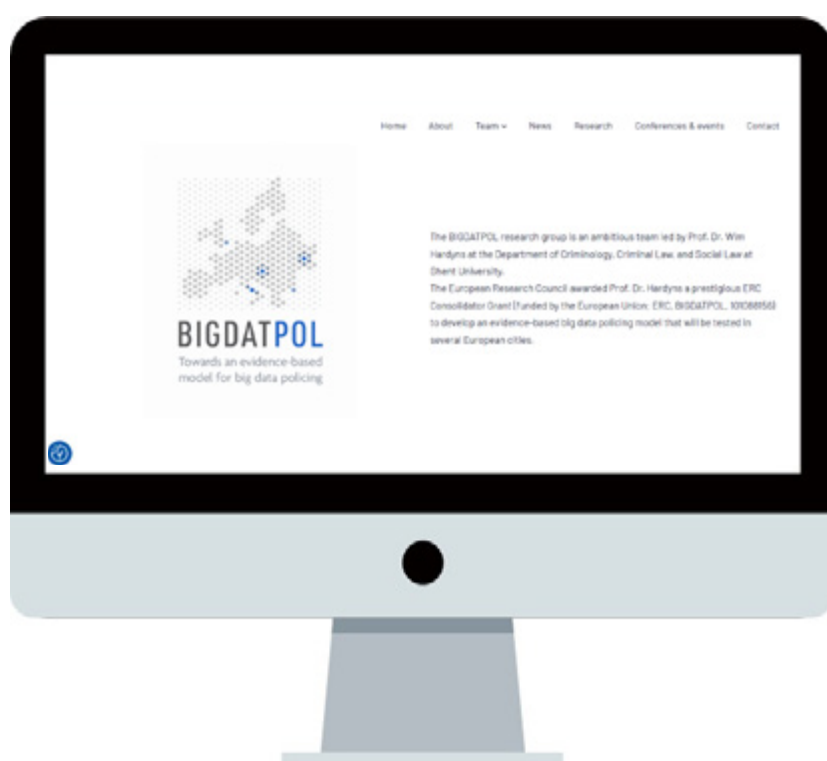
3. In the evening, during the **poster session**, our team was able to interact with interested and interesting researchers and answer questions about our research project.

We left with great insights and valuable connections and look forward to further refining our research!

Full programme on the [conference website](#).



WWW.BIGDATPOL.COM WEBSITE IS LIVE!



Our team is excited to introduce our new website www.bigdatpol.com!

We invite you to delve into our research on big data policing, explore our key findings and engage in meaningful discussions with us.

Meet the minds that make up our team and discover our mission and long-term research goals. Keep up to date with our latest activities by reading our news and updates.



The site also features an extensive archive of past presentations, peer-reviewed publications and media coverage highlighting our contributions to the field.

Don't miss our calendar of upcoming events, including conferences, webinars and workshops that you can attend or follow.

We will soon be launching a Q&A section where we will answer your most pressing questions and provide a platform for ongoing dialogue.

Looking further ahead, we will be blogging about emerging trends and challenges in big data policing, providing expert perspectives on the latest developments.

We look forward to sharing our journey with you, and building a dynamic community around big data policing.



THE HAGUE EVIDENCE BASED POLICING CONFERENCE



26 SEPTEMBER 2024

EVIDENCE-BASED POLICING IN DE PRAKTIJK

DEN HAAG

nsCr

SEBP

On 26 September, Dr. Thom Snaphaan (postdoctoral researcher BIGDATPOL) and Mark Crispel (Chief Constable, Local Police Zennevallei) provided a joint presentation at the [Evidence Based Policing conference](#) in The Hague, the Netherlands.

During this presentation, they outlined how we are working toward an evidence-based model for big data policing within this project. In particular, this presentation focused on the results from the pilot test in the local police zone Zennevallei. Chief Constable Crispel explained this from the perspective of the local police. A valuable collaboration of scientific research and practice!

nsCr
Nederlands Studiecentrum
Criminaliteit en Rechtshandhaving

SEBP
Society of evidence-based policing
Nederland

WARSAW EUROPEAN CONFERENCE OF THE SOCIETY FOR BENEFIT-COST ANALYSIS (SBCA)



The conference, hosted by the University of Warsaw, Faculty of Economic Sciences, aims to promote and improve the theory and application of the tools of benefit-cost analysis; that is, to have both a research and policy focus and to provide a forum for scientific exchange.

Researcher Charlotte Vandenbrande presented her scoping review on economic analysis in crime prevention programmes. The scoping review presented the state of the art in economic evaluations of crime prevention programmes.

Important takeaways from the conference include the need for a discount rate and sensitivity analysis. The discount rate is a monetary valuation technique that takes into account the fact that benefits now are worth less in the future. A sensitivity analysis helps to assess the robustness of the results by exploring how changes in key assumptions or input variables affect the results. The results of the scoping review show that not many economic analyses of crime prevention programmes currently include sensitivity analysis.

The main findings of the scoping review also show that the most commonly used measure in the evaluation of crime prevention programmes is the benefit-cost ratio. In its simplest form, this equation is the benefit/cost ratio. Whenever the benefit-cost ratio is greater than 1, the programme is beneficial.

The benefits in this ratio are not easy to monetise, and many studies define the benefits as the crime averted times the cost of crime. The latter is difficult to calculate and there are several approaches that can help in calculating the cost of crime. For example, some studies have calculated the cost of crime as the total cost of justice budgets, police budgets, court budgets, etc.

Another method is 'willingness to pay', where the cost of crime is defined after asking people how much they would be willing to pay to reduce a certain type of crime by a certain amount (e.g. how much would you be willing to pay to reduce burglary by 10%?).

The presented scoping review can be seen as a starting point for choosing an economic approach to evaluating big data policing.



CALL FOR QUESTIONS

We invite you to share your thoughts!

If you have any questions, concerns, or suggestions for improvement regarding experiences, challenges, specific conditions, or the effectiveness and efficiency of big data policing, please contact our team at bigdatapol@ugent.be or via [LinkedIn](#).

Your input will contribute to our research, and we will feature a Q&A section on our website to address these topics.

CALENDAR

24 OCTOBER KEYNOTE RADILAB, GHENT

Wim Hardyns will deliver a keynote at the study day organised by G4S, CPS and RadiLab on Radicalisation and Technology. The keynote will focus on violent extremism and the opportunities of big data analytics in this field. [More info.](#)

13 NOVEMBER ONLINE LECTURE NATIONAL SECURITY DAYS SLOVENIA

Wim Hardyns will give an online presentation on 'Data and security: an introduction to the BIGDATPOL research programme. [More info.](#)

28 NOVEMBER EXPERT COURSE DATA & AI, GHENT

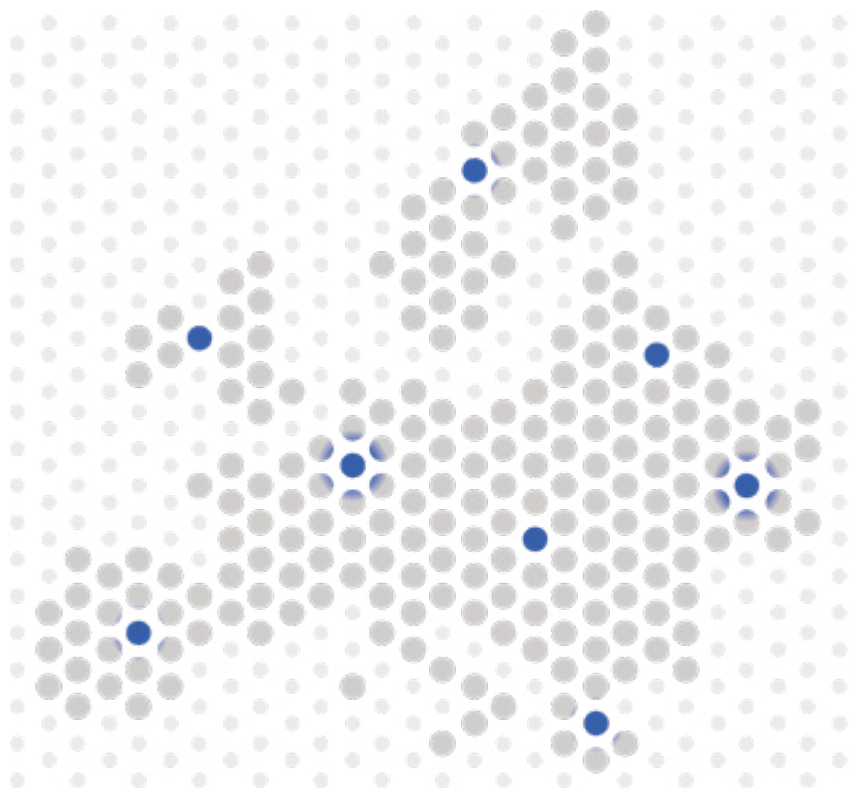
Interactive sessions and case studies delivered by the BIGDATPOL team will teach participants how AI and data analytics can be used to predict, understand and combat criminal activity. The expert course not only highlights the technological aspects, but also addresses the ethical and legal considerations involved in using AI in the security sector.

10 DECEMBER ODISSEI CONFERENCE, UTRECHT

Researchers Naomi Theinert and Robin Khalfa will present the paper 'Explaining predictions of crime in space and time – the contributions of Explainable Artificial Intelligence (XAI). [More info.](#)

13 DECEMBER STUDY DAY BIGDATPOL, ANTWERP

This study day, organised by the CPL, a professional network of executives within the Belgian police, will take place at the Provinciehuis in Antwerp.



-  bigdatpol@UGent.be
-  +32 9 264 84 99
-  Campus Aula
Universiteitstraat 4
9000 Ghent (Belgium)
-  www.bigdatpol.com
-  linkedin.com/company/bigdatpol

BIGDATPOL

Stay in touch

Funded by the European Union (ERC, BIGDATPOL, 101088156). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.



European Research Council
Established by the European Commission