What about AI in criminal intelligence? From predictive policing to AI perspectives

Patrick Perrot
Gendarmerie Nationale, Ministry of Interior, Paris, France

Abstract

Predictive policing is more and more developed around the world. TV-shows and fictions such as 'the minority report' or 'Person of Interest' spread a pre-crime effect that is, nevertheless, very different from reality. Many law enforcement bodies develop predictive analysis to find new opportunities against crime and it is generally dedicated to patrols. The Gendarmerie Nationale in France carried out through the concept of criminal intelligence a way to provide relevant information to describe, understand and foresee crime at different scales: operational, tactic and strategic. The aim is to upgrade the process of decision-making. Because crime is nor a random process neither a deterministic process, some features exist to characterise it. Obviously, it is very difficult and probably not possible to identify all features linked to crime evolution or criminal behaviour. Nevertheless, some characteristics are not so complicated to model in a formal mathematical structure. So, in the age of Big Data, applications of predictive analysis can be overtaken by artificial intelligence (AI). It is very developed in fields like medicine, finance or transportation and could on one hand provide new perspectives to fight crime but also on the other hand raise questions for future. Who will be the next organisation able to assure the best way to anticipate crime and criminal behaviour? AI could be defined as the capacity of a computer to model human reasoning. A grand challenge is opened for law enforcement but only if they are able to adapt their way of working to this new era. The scope of this paper is to describe the French development in predictive analysis and to open the potential use of artificial intelligence in different area of criminal intelligence without avoiding the risk of its new development.

***Key words*:**
Artificial intelligence, crime analysis, predictive policing, GAFAM, law enforcement

Introduction

As proposed by the International Association of Crime Analysts (2014) 'intelligence' is perhaps one of the most confounding terms in crime analysis. Indeed, it is very difficult to embrace all the characteristics of crime in one field. Nevertheless, criminal intelligence in France is considered as the way to collect any kind of information required to fight crime by anticipation. For several years, the fight against crime in France, based on a posteriori methods has not been any more efficient and able to contain the criminal evolution. Indeed, the possibilities available to offenders in term of mobility or IT resources, allow him to commit multiple crimes with a great effect and minimal risk. The techniques of investigation are until today essentially reactive as illustrated by forensic sciences. Since Edmond Locard or Alphonse Bertillon, who were two pioneers in forensics, science has contributed to the resolution of many crimes. However, an a posteriori point of view is not sufficient today to face new forms of offences and offenders. The appropriation of new technologies in the field of common crime as serious and organised crime requires an evolution of law enforcement to adapt their modus operandi to anticipation and proaction. This is the reason why the Gendarmerie Nationale develops a criminal intelligence integrating new methods of prediction.

Criminal intelligence is defined as a concept that regards investigation and public safety. While in the first case, the aim is to analyse individuals or groups involved in crime, in the second case, the aim is more preventive by providing a macro view of crime and evaluating explaining features at different scales (short, medium and long term). In crime investigation as in public safety, criminal intelligence consists in collecting, analysing and valorising data to propose a way to reduce or decrease the number of offences by anticipation. Anticipative policing appears as an application of mathematical techniques to identify likely targets and prevent crime or solve past crimes by making statistical predictions. For 2 years, the Gendarmerie Nationale has developed a predictive approach of criminal risk (Berk, 2012; Perrot 2014; Perrot and Achi, 2015) oriented to common crime as serious and organised crime. Data analysis is, of course not, a new concept but the process of examining large data sets containing a variety of data to uncover hidden patterns, unknown correlations, crime trends, criminal preferences and other useful information needs to be encouraged in the Big Data era. Mathematicians are able to predict with a reasonable doubt hotspots, as well as when each type of crime might occur, by a careful and probabilistic analysis. But, within the development of a discipline that finds a name in 1956, Artificial Intelligence (AI), exciting new opportunities were now available. Is it reasonable to think that artificial intelligence could help law enforcement in the fight against crime? This paper will discuss the appearance of AI in many applications since 1956, its applications against crime and the interest for law enforcement to use it in criminal intelligence.

What about AI?

In the field of AI, Alan Turing appears as a pioneer. In 1950, he defined a test based on indistinguishability from undeniably intelligent entities-human beings and designed to provide a satisfactory operational definition of intelligence (Turing, 1950). The computer passes the test if a human interrogator, after posing some written questions, cannot tell whether the written responses come from a person or not. In 1950, the term 'AI' didn't exist but the idea that machines will be able to imitate human reasoning had emerged. Another great mathematician, John McCarthy, proposed in a summer school (Dartmouth College conference) in 1956 a name (or two letters) to a concept: Artificial intelligence (AI) (McCarthy, 1959). McCarthy tried to develop a language able to translate human reasoning in computer instructions. At this time, it was very difficult to imagine all the possibilities of such a way, but today in the Big Data era, many perspectives are opened and a great challenge is in front of law enforcement. AI is based on mathematical models used to solve real-world problems and have demonstrated all its relevance in many domains as varied as biometrics, medicine, transportation, or education.

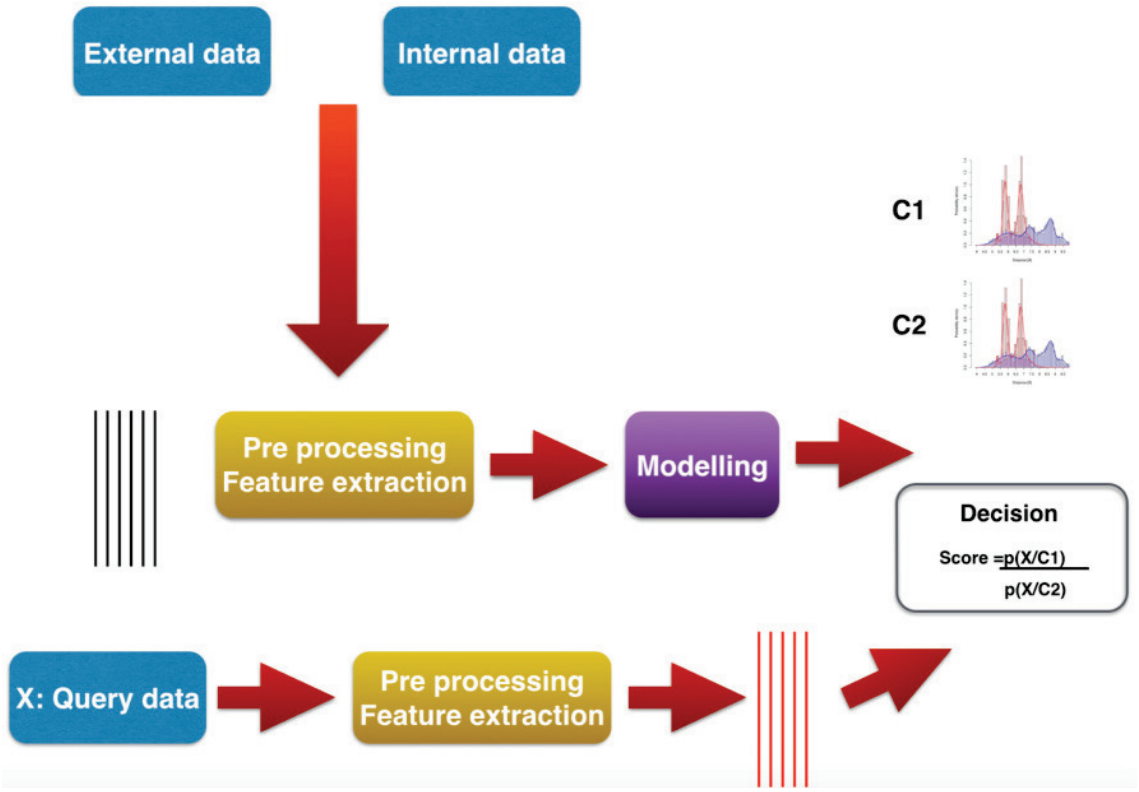
The AI core is called machine learning (Breiman, 2001a; Marsland, 2009) and its extension deep learning or q-learning for tomorrow. This discipline aims to understand the fundamental principles of learning as a computational process and combines tools from computer science and statistics from large datasets. Machine learning is very relevant to recognise but also to predict new patterns and provide a possibility to improve performance with feedback. To this end, machine learning has been applied, ranging from fuzzy logic reasoning to artificial neural network creation, in an effort to imitate the human logic and the way correlations or inference may be achieved. It is based on a set of observations to learn a model like in the case of the famous Pavlov's Dog experiment. In other words, within an expert system based on machine learning, we have a training sample of n observations on a class variable 'Y' that takes values (1, 2,..., k,) and p predictor variables, (X1, ..., Xp). Our aim is to find a model for predicting the values of 'Y' from new 'X' values.

Among the large range of machine learning methods, we can cite neural networks, Gaussian mixture models, classification tree or random forest and so on. Neural networks are based on modeling the neurons and feeding the network a set of training data to find patterns. A Gaussian mixture model is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters. Classification tree consists in modeling by recursively partitioning the data space and fitting a simple prediction model within each partition. As a result, the partitioning can be represented graphically as a decision tree. Based on the classification trees theory, random forests (Breiman 2001b) try to classify a new object from an input vector, put the input vector down each of the trees in the forest. Each tree gives a classification, and we say the tree 'votes' for that class. The forest chooses the classification having the most votes (over all the trees in the forest).

So, the machine learning theory is based on different phases: feature extraction, training, test and evaluation. Feature extraction is a type of dimensionality reduction that efficiently represents interesting parts of a query object as a feature vector. In the training phase, the features of an object or pattern are stored as reference features to generate numerical templates for future comparisons. The numbers of reference templates that are required for efficient recognition depend upon the kind of features or techniques that the system uses for recognising the object. In the recognition phase, features similar to the ones that are used in the reference template are extracted from an input object whose identity is required to be determined. The recognition decision depends upon the computed distance between the reference template and the template devised from the input utterance. The more training data, the better the pattern recognition. Figure 1 illustrates an expert system based on a phase of enrolment of data to create model, a phase of pattern recognition and a phase of decision.

The level of performance of an expert system is quantified most typically by a 'receiver operating characteristic', called 'ROC curve' or a detection error tradeoff curve called DET curve. This curve reveals the compromise between the 'false acceptance rate' and the 'false rejection rate'. The false acceptance rate is the frequency with which query data from different sources are erroneously assessed to be from the same source and the false non-match rate is the frequency with which query data from the same source are erroneously assessed

Figure 1 — Expert system principle



to be from different sources. The performance of a system falls on a point on the ROC curve whose location is a function of the matching 'threshold' applied. A higher match threshold reduces false acceptance rate and increases false rejection rate. On the contrary, a lower match threshold reduces the false rejection rate but increases false acceptance rate.

The decision step is a binary hypothesis testing problem expressed by:

H_0: impostor object

H_1: client (real) object

The $P_{fa}=P(x=1|H_0)$ is the probability of false acceptance then $P_{fr}=P(x=1|H_1)$ is the probability of false rejection

According to the Bayesian theory, these two kinds of errors are weighted by costs and summed into a single cost function, the Bayesian risk function:

$$B_{risk}= P(H_0).C_{fa}.P_{fa}+P(H_1).C_{fr}.P_{fr} \quad (1)$$

with $P(H_0)$: probability of an impostor object, $P(H_1)$: probability of a genuine object, C_{fa}, cost of false acceptance, C_{fr}, cost of false rejection.

$P(x|H_0)$ is compared to a threshold that divides the decision region between a region of acceptance and a region of rejection. If an object's matching score happens to fall above the thresholds, it is considered as genuine, if it is below as imposter.

One of the main advantages of an artificial intelligence system is its capacity to continuously learn any kind of information increasing the efficiency of a decision. Such a system takes into account different views, different perspectives and thus is proposing a more complete analysis. But what is important to consider is that AI is first an empirical science. AI follows a hypothesis-and-test research paradigm. The performance of these systems are very linked to the databases and to the algorithm used.

AI has developed so rapidly over the past few years struggling lesser to make sense of what they see or hear. Computers can now outperform humans in some cases. This is the case in the field of object recognition, face recognition, facial expression, speaker recognition and even emotion identification.

So, AI applications are more and more developed and common in our world. In 1997, Deep Blue, an expert system built by IBM, has already defeated the world chess champion, Garry Kasparov, in a six-game match after a defeat in 1996. Then, in 2011, another system from IBM called Watson, in reference to John Watson, the founder of the society, won in a jeopardy

game against humans. And recently, AlphaGo from Google, based on Monte-Carlo tree search with deep neural networks defeated a human European champion. Independently from this cases revealed by media, AI progresses day after day at a very high speed in many areas. For instance, Google, Apple, and Microsoft are competing to transform vehicle transport with self-driving vehicles. Some applications open up new ways for human to interact with embodied conversational agents (Wiendl et al., 2007) able to perceive not only the virtual but also the physical world as well as reactive behaviour control. Google and Facebook are designing chatbots that make decisions for users about diverse activities like commercial, shopping or travel arrangements. Such innovations could not be external to police affairs.

Application against crime: From predictive analysis to AI

The first step before the development of AI could be the predictive analysis development as we see in many countries. Among concrete applications, we can cite the modeling of crime characteristics in order to anticipate new occurrences. The Washington Times headlined about prediction: 'Never a crystal ball when you need one'. Even if such a ball does not appear as the good way, forecasting based on scientific methods is a real way of progress for criminal intelligence. Spatial and temporal methods appear as a very good opportunity to model criminal acts. Common sense reasoning about time and space is fundamental to understand crime activities and to predict some new occurrences. The principle is to take advantage of the past acknowledgment to understand the present and explore the future. Based on multiple methods including exponential smoothing algorithms (simple, double, triple), autoregressive integrated moving average techniques and neural networks, a prediction analysis is carried out on different offences. These models are fitted to offence time series data either to better understand the data or to predict future points in the series. They are applied in some cases where data show evidence of non-stationarity. Results are derived from two different sets of past data. The first one is used to train the algorithm and build the model and the second one is used to evaluate the performance of the model. Data from 2008 to 2013 constitutes the training set and data from 2014, the evaluation set. Based on the prediction model, a future evolution is proposed for 2015 as illustrated in Figures 2 and 3. The curves (below) illustrate a temporal evolution of a specific offence month per month. Blue colour characterises the real evolution, red, the prediction model and green, the predictive evolution.

Based on this curve, it is possible to optimise the allocation of resources during some specific periods but also to evaluate the performance of some modus operandi used in the past (last year or last month). Indeed, the delta between real and predicted value in 2014 must be explained in order to evaluate the efficiency of police operations and to optimise next actions. This kind of curve is developed for many offences and so, a prioritisation of actions

Figure 2 — Construction of the prediction model

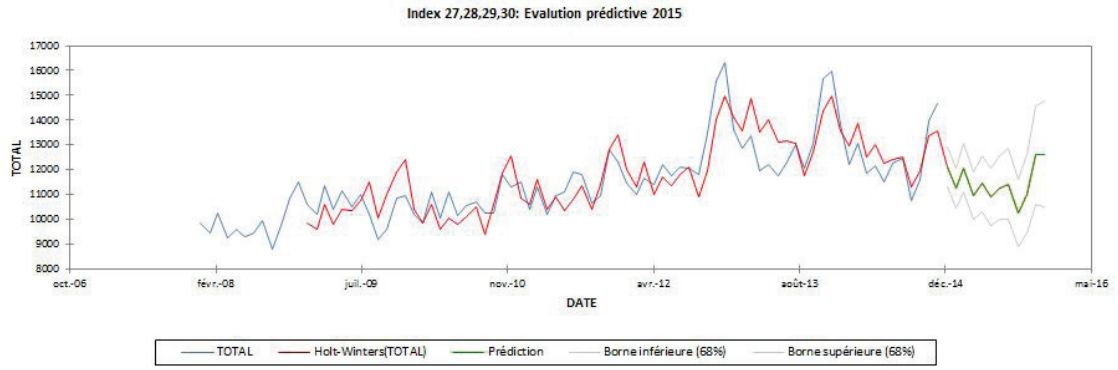
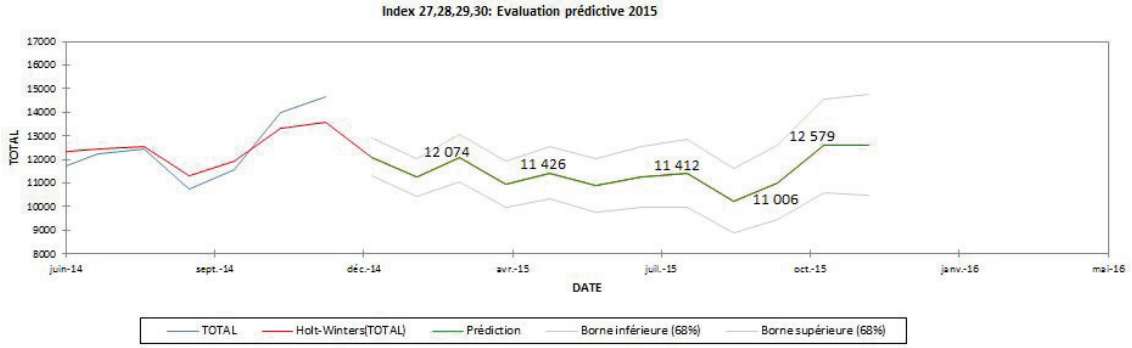


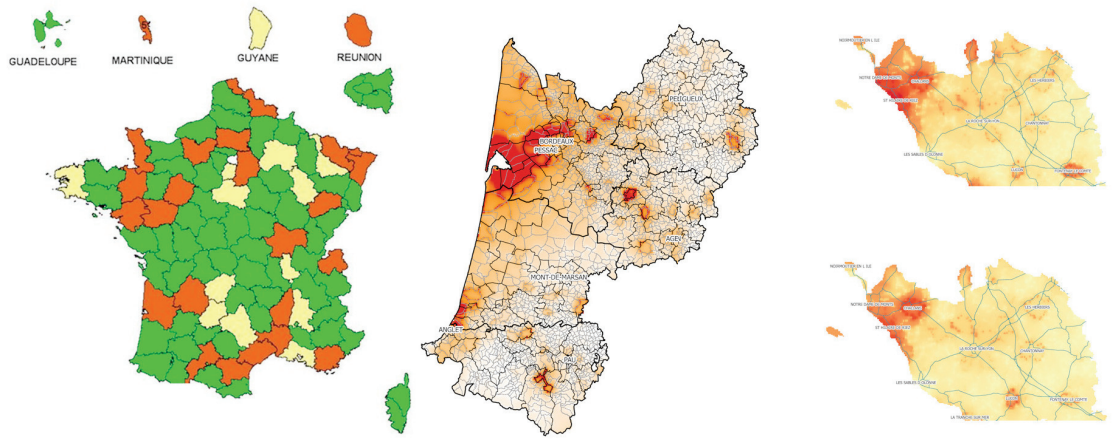
Figure 3 — Predictive evolution



can be proposed targeting the right offence at the right time. Complementary to a temporal view, a spatial analysis can also be carried out at different scales as proposed in Figure 4.

The main interest is to identify trends, patterns, or relationships among data, which can then be used to develop a predictive model and propose short, medium and long-term trends (Hoaglin et al., 1985) in order to inform police service at different levels. A map visualisation is very interesting and relevant to anticipate crime or criminal moving by evaluating the places of concentration or the dispersion movement. Such work can be completed by the association of external data in order to find explanation. For instance, based on a regression model it is possible to explain crimes like burglaries from social and economic data like urban development and population growth.

Figure 4 — National-Regional-Departmental view of a predicted offence



Because crime is neither a random nor a deterministic process, some features exist to characterise crime and perhaps offenders or police officers. Based on this assumption, it is possible to mechanise some tasks and upgrade predictive policing by applying AI techniques on the question of crime and to get benefits from machine intelligence.

Web 3.0, also considered as the semantic web, is a space where the data has its own meaning and its own means of production from connected objects. This web offers the possibility to deliver observational, behavioural and tailored content to individuals rather than to 'crowds'. Web 3.0 is a web of targeting based on massive data. From this concept a new form of police could emerge: a police 3.0 or 4.0 taking account of the analysis of massive data, the exploitation of connected devices and a capacity to provide individual profiles by anticipation. The question is to know who will be the master of this new police? The answer to this question is a real subject and challenge for the next years. The GAFAM (Google, Apple, Facebook, Amazon, Microsoft) and the NATU (Netflix, Air BNB, Telsa, Uber) global corporates collect more and more data every day and have a real capacity of analysis and produce objective results. In 2015, Bill Gates expressed that AI is entering a period of rapid advances. AI will fundamentally change how humans move, communicate and live. Today personal data and private information cannot be fully controlled. GAFAM are at the forefront of innovation in artificial intelligence, with active research exploring virtually all aspects of machine learning, including deep learning and more classical algorithms. They gather large volumes of direct or indirect evidence of relationships of interest, applying learning algorithms to understand and generalise. In the field of crime analysis it is easy to imagine some concrete applications:

— To recognise a known criminal in a specific area and send an email on a personal smartphone;

— To identify geographical and time hotspot areas of crime;

— To make a profile of criminal based on massive data;

— To indicate the level of multiple offences in a specific area; and

— Why not to replace a police officer by a virtual agent in specific tasks.

It could be very exciting to visualise on one's own smartphone a risk of theft or aggression on a specific area or to get an estimation of the number of pickpockets around us. It might provide the citizen with a sensation of control of his or her own security — but isn't it an illusion? Indeed, all these applications cause a risk for privacy and for the power to decide. In addition, without any control of the data, it will be very difficult to evaluate the reliability of the information. The risk is real and the best way to protect people from abuses and to avoid a police driven by ROI (Return On Investment), is to allow and rely on the development of AI applications by law enforcement.

In an era of accountability, law enforcement cannot rest on past accomplishments against crime for very long. Law enforcement must go one step further but step by step by rigorously respecting privacy. It is more and more important to be flexible and creative to face a criminality in constant and quick evolution. This new vision for future is a great and exciting challenge. Using computer to analyse how offenders react to questions combined with the ability to identify what sort of people they are, could also provide new opportunities to help investigators.

An expert system is able to autonomously learn crime activities and behaviours which otherwise would be masked in a global environment. From a theoretical point of view, AI can be used in three different cases:

— To model criminal acts;

— To model behaviour and criminal way of reasoning;

— To model behaviour and investigator way of reasoning.

The objective is to extract knowledge from these three sources and why not from a fusion of these sources. Nevertheless, to transform a theoretical point of view to practical applications is not so easy.

A possible use of AI is to model specific profiles of criminals. The principle of this approach consists in evaluating the possibilities that a suspect relates to an a priori class. The advantage of an AI is to train the model from criminological theory and from real case reports. These kind of applications could be realised to build a class model for specific criminals but also for victims. Analysis of the patterns formed by prolific offenders could be built on many elements, like their movements, their area of living, advertising or working, their habits, their type of crime, their previous convictions, their home, their daily activities, their social networks, the offence locations, etc.

Based on the same principle and in the case of financial crimes linked to sensible companies, a profile based on social engineering could define an evaluation risk of attacks. A first condition is to get a history of past cases in order to build a dataset of victims and another one of non-victims. The competitive hypothesis developed in the decision process is:

H_0: non-victim company

versus

H_1: victim company

and the risk for a query company to be victim is calculated by (1).

This kind of analysis starts to be used in order to find the most probable possibility. In future, it could also be possible to model the behaviour and the investigator's reasoning. But it is necessary to know: what is an investigator? — and this is a very complex problem to solve. In many cases, the investigation process is a logical enterprise in a logical environment, formed by the legal procedure. In addition, an investigator uses his own experience to increase his relevance. These different aspects can be modeled by an expert system based on the principle of training. One source of the investigator reasoning is the results of interviews carried out with the agents that could be used to train a model. A police officer on patrol most likely uses deductive reasoning and learns everything by experience (Bosio, 2011). So, the challenge for an expert system is to be able to incorporate experience and a way of reasoning. Yet, knowledge and intuition of the police officer play a central role. All the process is not logical and an investigator in front of a situation needs to keep an open and adaptive attitude. Technical and logical knowledge, although necessary, is not sufficient to account for the global process of investigation. Patrizio Bosio emphasises that the everyday experience of a police officer is imbued with grey areas and with excellent, albeit concealed, knowledge.

Currently, we can consider that a virtual agent able to provide objective help to a real investigator doesn't seem realistic because of the heterogeneity and the complexity of the situation that is not uniquely logical. Formal or informal perception plays an important role in the grip on reality for investigators. Because the human brain is not a chess program, AI is not completely ready today to emulate it. Understanding criminal investigations also requires inferring a hidden factor, namely, the intention of the police officer. But we cannot exclude for the future that the extension of AI in this field is based on an analysis of police officer patterns. Analyses could for instance include experience, age of investigators, trajectories, modus operandi of investigations, crime type, and so on.

In conclusion: the integration of AI is still a process under construction. Machines are learning to see in increasingly reliable and useful ways, opening up a wide range of opportunities and perspectives for law enforcement. AI can increase the capacity to receive real-time alerts of abnormal behavior and quickly respond to time-sensitive and critical events. Indecision and delays are the parents of failure, the aim is to upgrade human decision-making thanks to AI. The risk is to see these perspectives developed by private societies or industrial groups instead of law enforcement.

Conclusion

The aim of this paper is to prepare law enforcement to the unavoidable use of AI techniques. Generally in AI, systems are autonomous and can decide what to do and then do it. In the case of security, these kind of perspectives are not admissible, the decision must be human. The human decision-maker must be considered as the centrepiece for police operations even if powered by AI that accelerates the decision. Human decisions must be a sanctuary in the field of police activities. In such a way, AI methods provide mainly a support in the process of decision-making. It is an exciting time for the field, as connections to many other areas are being discovered and explored, and as new machine learning applications bring new questions to be modeled and studied. It is safe to say that the potential of AI and its application against crime lie beyond the frontiers of our imagination but must be limited by questions of privacy. Information communications technologies cannot supplant privacy obligations. This is one of the main reasons why law enforcement must be engaged in this way of development. The grand challenge is not for tomorrow but for today. And don't forget that AI can also be used by the criminal offender. What will be the future criminal uses of automated crime? How will criminals take advantage of advances in AI to extend and improve their criminal activities? We have already seen the very beginning of these activities in cyberspace. But what happens when computers think and can improve themselves on new forms of previously unanticipated forms of criminality?

References

- Berk, R. A. (2012), *Criminal justice forecasts of risk: a machine learning approach*, Springer.
- Breiman, L. (2001a), Statistical modeling: two cultures (with discussion). *Statistical Science*, 16, 199-231.
- Breiman, L. (2001b), Random forests*, Machine Learning*, 45, 5-32.
- Bosio, P. (2011), *Knowledge from experience of a police officer: a grounded study*, *CEPOL European Police Science and Research Bulletin*, (4), 12.
- Hoaglin, D. C., Mosteller, F. and Tukey, J. (1985), *Exploring data tables, trends, and shapes*, John Wiley.
- International Association of Crime Analysts (2014), *Definition and Types of Crime Analysis*, Standards, Methods & Technology (SMT), Committee White Paper.
- Marsland, S. (2009), *Machine learning: an algorithmic perspective*, CRC Press.
- McCarthy, J. (1959), Programs with Common Sense, *Proceedings of the Teddington Conference on the Mechanization of Thought Processes*, 756-91.
- Perrot, P. (2014), L'analyse du risque criminel: l'émergence d'une nouvelle approche, *Revue de l'Electricité et de l'Electronique*.
- Perrot, P. and Achi, K. T. (2015), Forecasting analysis in a criminal intelligence context, *Proceedings of the International Crime and Intelligence Analysis Conference*.
- Turing, A. (1950), Computing Machinery and Intelligence*, Mind*, 59(236), 433-460.
- Wiendl, V., Dorfmüller-Ulhaas, K., Schulz, N. and André, E. (2007), Integrating a virtual agent into a real world, *Proceedings of the 7th international conference on Intelligent Virtual Agents*, 211-224.