Reconciling Adapted Psychological Profiling with the New European Data Protection Legislation

Keeley Crockett[1[0000-0003-1941-6201]] and Jonathan Stoklas [2] and James O'Shea[1[0000-0003-1941-6201]] and Tina Krügel [2] and Wasiq Khan[1[0000-0002-7511-3873]]

[1] School of Computing, Mathematics and Digital Technology, Manchester Metropolitan University, UK
[2] Institute for Legal Informatics, Leibniz Universität Hannove, Hannover, Germany
{k.crockett,j.d.oshea}@mmu.ac.uk, jonathan.stoklas, kruegel @ iri.uni-hannover.de}

Abstract. Adaptive Psychological Profiling systems use artificial intelligence algorithms to analyze a person's non-verbal behavior in order to determine a specific mental state such as deception. One such system known as, Silent Talker, combines image processing and artificial neural networks to classify multiple non-verbal signals mainly from the face during a verbal exchange i.e. interview, to produce an accurate and comprehensive time-based profile of a subject's psychological state. Artificial neural networks are typically black-box algorithms; hence, it is difficult to understand how the classification of a person's behaviour is obtained. The new European Data Protection Legislation (GDPR), states that individuals who are automatically profiled, have the right to an explanation of how the "machine" reached its decision and receive meaningful information on the logic involved in how that decision was reached. This is practically difficult from a technical perspective, whereas from a legal point of view, it remains unclear whether this is sufficient to safeguard the data subject's rights. This chapter is an extended version of a previous published paper in IJCCI 2019 [35] which examines the new European Data Protection Legislation and how it impacts on an application of psychological profiling within an Automated Deception Detection System (ADDS) which is one component of a smart border control system known as iBorderCtrl. ADDS detects deception through an avatar border guard interview, during a participants' pre-registration, to demonstrate the challenges faced in trying to obtain explainable decisions from models derived through computational intelligence techniques. The chapter concludes by examining the future of explainable decision making through proposing a new Hierarchy of Explainability and Empowerment that allows information and decision-making complexity to be explained at different levels depending on a person's abilities.

Keywords: Psychological Profiling, Deception Detection Artificial Neural Networks, Decision Trees, GDPR.

1 Introduction

Psychological Profiling is a technique best known as a tool used within criminal investigations utilising methodologies from both law enforcement and psychology [1]. It involves the detailed and intricate analyses of the non-verbal behaviour of a person, often in an interview situation, to detect their mental state. The expertise and training required by a human to undertake this kind of profiling is complex – requiring simultaneous conjecture of many non-verbal signals. Adaptive psychological profiling utilises computational intelligence techniques to build models of non-verbal behaviour for different mental states, i.e. deceptive behaviour or more recently to detect comprehension levels in education. For example, Silent Talker (ST) [2], a profiling system for lie detection, uses hierarchies of neural networks to module deceptive behaviour. However, neural networks are by nature 'black boxes' in which it is difficult to understand how the trained networks determine if a person is deceiving or not.

The European data protection reform package that is applicable since May 2018 consists of the General Data Protection Regulation (Regulation 679/2016/EU, "GDPR") and the "Law Enforcement Directive (680/2016/EU). The GDPR potentially has a worldwide impact on business models and research activities carried out within industry and academic institutions that utilise computational intelligence (CI) algorithms with respect to personal data [3]. Specifically, it states the rights of an individual not to be subject to automated decision-making, such as profiling, unless it is (a) necessary for entering into or performance of a contract between the data subject and a data controller, (b) authorised by Member State's laws, or (c) it is based on the explicit consent of the data subject. In addition, in any aspect of automated decision-making, the individual has the right to human intervention (opt out) as well as to be provided with an explanation of how the automated decision was reached. This would be achieved through disclosure of "the logic involved" (article 13, GDPR, [4]). When profiling, the data controller should use appropriate mathematical and statistical procedures and data should be accurate (up to date and free from bias) in order to minimize the risk of errors. This legislation presents many challenges when using CI for modelling complex problems that involve people. How do we provide an explainable decision suitable for all stakeholders when using 'black box' CI algorithms? The stakeholders are the experts who designed, validated and tested the system, the business or customer who commission the system and the data subject who receives the automated decision from the system. This chapter explores this issue using an application of an automated deception detection system (ADDS) utilised within a pilot system known as iBorderCtrl, which detects deception through an avatar border guard interview during a participant's pre-registration. The final stage of the deception detection architecture is a single neural network classifier. In section 5 experiments, this is re-placed with a traditional decision tree to provide a set of rules on how decisions about deceptive behaviour are reached. The complexity and size of the rule sets produced show, that whilst an expert, may have some understanding of the rules, it would be extremely difficult for a member of the public to understand and even the expert could at present not be able to say precisely why these particular rules were derived or explain what they mean.

Section 2 of this chapter defines what is meant by psychological profiling in the context of this work, whilst Section 3 explains the legal perspective of some aspects of automated decision making in light of the GDPR. The case study of profiling EU travellers is described in Section 4 and used to illustrate the challenges of developing explainable profiling systems. Section 5 provides the methodology used to conduct empirical experiments on a deception detection profiling system and presents results using both neural networks and decision trees in terms of explainability. This section also considers the future of explainable decision making in terms of a proposed Hierarchy of Explainability and Empowerment. Finally, section 6 provides some important considerations for both the legal and computational intelligence communities.

2 PSYCHOLOGICAL PROFILING

Adaptive Psychological Profiling is the process of determining a person's internal mental state (beliefs, desires, and intentions) through analysing their external behaviour by means of Computational Intelligence (CI) based components. Furthermore, it is based on a generic architecture, which is adapted to different application domains and optimised through a process of machine-learning. The first such architecture is known as "Silent Talker" (ST) [2, 5]. ST uses complex interactions between non-verbal features in a moving video feed from an interviewee to classify truthful or deceptive behaviour. The ST architecture has been adapted for different internal mental states. One such adaptation is for comprehension in intelligent tutoring, in the classroom [6]. Another ethnic / cultural adaptation extended comprehension classification to Tanzanian women for informed consent in a clinical trial [7]. Other ongoing work includes an avatar based deception detection interview integrated into a smart border crossing system [8, 9]. The case study used in this chapter focuses on the complex problem of the psychological profiling of deceivers – the next section looks more deeply into the science of lying and why this in particular is a challenging problem for computational intelligence in terms of building a model and in trying to explain automated decision making.

2.1 The Science of Lying

There are various different types of lie, with different contextual motivations and different ways of classifying them. For example, Ganis et al. [10] used two classes, whether the lies fit into a coherent story and whether they were previously memorized. Alternatively, Feldman et al. [11] presented a taxonomy of lies with 10 codings, for lies produced by participants with three different self-presentational goals. Regardless of context, there is a general psychological principle that the act of deceiving produces changes in behaviour which has a long history dating back to the Hindu Dharmasastra of Gautama, (900 – 600 BC) and the philosopher Diogenes (412 – 323 BC) according to Trovillo [12].

There are a number of factors, proposed by psychologists, which may be influential drivers of behavioural change during deception. These include general arousal / stress, cognitive load, behaviour control and special cases of arousal, guilty knowledge and duping delight. Stress is the oldest driver to be measured for lie detection. Following work by Angelo Mosso in the late 19th and early 20th centuries using pulse and blood pressure, the polygraph was invented by Larson in 1921 (International League of Polygraph Examiners, [13]). The Cognitive Load driver derives from the work of George A. Miller [14], whose Magical Number 7 (+/- 2) indicated that there were a limited number of “mental variables” that an individual could process concurrently. Therefore, someone trying to construct and remain consistent with a false account would be under increased cognitive load. Behaviour control occurs when deceptive interviewees deliberately try to control themselves in order to make an honest and convincing impression. It is postulated that attempts to control behaviour will increase in higher-stakes scenarios [15]. Guilty knowledge (Concealed Knowledge) is a test of whether a suspect has information related to a crime that an innocent person would not possess. When exposed to such information an interviewee is expected to produce a reaction detected by instrumentation [16]. Duping delight is believed to occur in an interview when the deceiver experiences pleasurable excitement at the prospect of successfully deceiving the interviewer, particularly in the presence of observers [17].

2.2. Automated Lie Detection

The field of computational intelligence provides a wealth of algorithms which are suitable to build models of deceivers automatically. Silent Talker (ST) [2, 5, and 18] differs from many other lie detectors in its assumption that deceptive non-verbal behaviour is the outcome of a combination of psychological drivers and that it cannot be characterized by a simple, single indicator. ST uses complex interactions between multiple channels of microgestures over time to determine whether the behaviour is truthful or deceptive. A microgesture is a very fine-grained non-verbal gesture, such as the right-eye moving from half-open to closed. This can combine with other microgestures from the right eye to detect a wink or both eyes to detect a blink. Measured over time these can combine to measure blink rate. Complex combinations and interactions of (typically) 38 channels and interactions between them can be compiled into a long vector, over a time slot, which can be used to classify behaviour as truthful or deceptive over the slot. Microgestures are significantly different from micro expressions (proposed in other systems), because they much more fine-grained and require no functional psychological model of why the behaviour has taken place. Furthermore, because there are so many channels contributing to the analysis, behaviour control is infeasible. Typically, using a recording device such as a web cam, salient features (e.g. an eye) are identified in an individual video frame by a layer of object locators. The states of the objects are detected by the pattern detectors (e.g. eye half open). The channel coders compiled the outputs of the pattern detectors over time (e.g. sequence of eye movement indicating a blink) and the deception classifier uses this long vector compiled by the channel coders. The ST approach to lie detection is based on a “black box” model, the conjecture that these and other (unknown) factors act as drivers of non-verbal behaviour, resulting in distinctive features that can be used to discriminate between deceivers and truth-tellers. Silent Talker is in itself an automated profiling system and is being piloted as a basis for an automated deception detection system to profile travellers crossing European borders at a pre-registration phase and will be described in section 4 of this chapter.

3 The GDPR

3.1 Automated decision-making under the GDPR

From a legal perspective, various issues arise. In 2016, the European Union agreed on a data protection reform package including the General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679, [4]), which is applicable since 25 May 2018. The GDPR introduces various new regulations which affect both profiling and the use of computational intelligence-based systems. Despite the fact that all general provisions have to be met, such as the data protection principles (art. 5 GDPR) and procession upon a legal basis (art. 6 GDPR), as explained above, for profiling and automated decision-making there is a specific provision in art. 22 GDPR. According to art. 22 (1) GDPR, an automated decision is a decision based solely on automated processing, including profiling of a person, which produces legal effects concerning him or her or similarly significantly affects him or her. One of the most obvious challenges in this regard is the fact that almost any decision in an increasingly digitized world might have at least a mediate legal effect as well [19]. Therefore, the interpretation of this requirement should be rather restrictive [20], whereas any single case shall be assessed based on objective factors [22]. While this result seems to be sound and necessary, persons without a legal background might face difficulties in assessing whether their decision is to be seen as automated decision-making or not. An automated lie detection system, however, will most probably always cause significant effects on the persons, both deriving from the possible use-cases, as well as with regard to personality rights in general (e.g. reputation).

3.2 Safeguards for automated decision-making

For decisions falling within the scope of art. 22 GDPR, certain safeguards need to be considered. Following the principles described in art. 5 GDPR, automated decision-making only can take place for specific purposes, and it must be necessary, legitimate and proportionate. Also, according to art. 22 (3) GDPR, the data controller shall “*implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision*.”

Consequently, it is not possible to apply automated decision-making without offering a secondary system which involves human beings. In that regard, automated decision-making would act as a filter, where only those cases which would result in negative consequences for the data subject would be checked by a human being. However, it is important to ensure that such a check by human beings is not biased by the previous automated results. From a legal point of view, it remains unclear whether decision assistance systems are being covered by art. 22 GDPR. According to the wording of art. 22 GDPR, it shall only apply to a decision based solely on automated processing, which does not include assistance systems [21]. In fact, art. 22 GDPR only ensures that human beings are not subject to a decision by a machine [22]. Therefore, at present assistance systems are in general not covered by the specific provisions on automated decision-making included in the GDPR. However, to qualify as human involvement, it therefore has to be ensured, that the involvement is carried out by someone who has both, the authority and competence to change the decision [20]. Nevertheless, it might be possible that human beings, even though they have the authority and competence will just follow the system's recommendation, in particular if the accuracy of such a system is extremely high. Proper training and sensitization (both with regard to the risks deriving from automated decision-making and knowledge on fundamental rights) is required. Also, organizational measures to ensure that human beings checking automated decisions do not have any negative consequences, such as the burden of proof or liability when deviating from the machine's decision, are required, as this would otherwise undermine the right to human intervention and would lead to a de-facto decision through the machine.

Apart from these issues raised by the safeguards explicitly mentioned in the GDPR, further issues can arise from the principle of non-discrimination. While art. 22 GDPR aims to ensure that the data subject does not face any negative consequences through a decision made by a machine, specific provisions on bias in algorithms cannot be found in the GDPR. Instead, general principles such as data accuracy (art. 5 lit. d) GDPR) and fundamental rights need to be considered. In the context of automated decision-making and fundamental rights, bias of algorithms is one of the most relevant issues. While from a technical point of view, a decision would be only wrong if the system was not working as intended, results can, although technically correct, nevertheless be wrong from an ethical and legal point of view. Biased algorithms, therefore, are rather a regulatory issue than a technical one. If an algorithm relies on biased data in any of the steps for development and use, results will be biased as well [23]. A common ground for this can be found in both disciplines, though: Data used for machine-learning and automated decision-making needs to be accurate, of good quality and shall not lead to discrimination. Therefore, on the input side, preventing and detecting inaccurate data is crucial for the application of artificial intelligence and machine-learning. For prevention, an impact assessment and the selection of data to learn from is a very important safeguard to ensure that the data subject's fundamental rights are not being violated; for the ongoing use on the output side, tools to constantly monitor decisions and identifying possible issues with bias is required [23]. As one of the most important steps to avoid bias is transparency [24], the “black box” phenomenon as described above is a specific challenge in that regard.

For the use of artificial intelligence in general, but also in the context of deception detection and border security, further issues can be identified, such as a violation of human dignity (i.e. caused by an increase in human-machine interaction with smart systems), for instance due to an objectification of the human being and a possible stigmatisation caused by false positives. It has to be noted, though, that these issues rather derive from other fundamental rights than the right to privacy, meaning that the GDPR alone is not sufficient to tackle the challenges for the legal framework raised by artificial intelligence.

Consequently, additional legislation might be required. However, whether these changes have to be addressed by a (fundamental) reform or rather lead to small adjustments depending on the use-cases in which artificial intelligence is applied [25], is still under discussion. Legislative approaches such as the strategy on artificial intelligence in Germany, which contains the establishment of a legal and ethical framework for research on and the use of AI [26], or the European parliament resolution containing recommendations to the Commission on Civil Law Rules on Robotics [27], or the Draft Ethics Guidelines for trustworthy AI [28], show that the legislator is increasingly aware of the issues, while precise regulations are not available yet. Also, the principle of proportionality requires all legislative approaches to consider if new technologies are actually beneficial for the society. This needs to be followed by an assessment on how ethical issues could be mitigated and the rights and freedoms of citizens can be safeguarded.

3.3 Information obligations and their extent

Additional obligations can be found in the data subject’s rights: According to art. 13 (2) lit. f), 14 (2) lit. g) and 15 (1) lit. h) GDPR, the data controller is required to inform the data subject about:

- the existence of automated decision-making as referred to in art. 22,
- meaningful information about the logic involved, and
- the significance and the envisaged consequences of such processing for the data subject.
- In addition, information has to be provided in concise, transparent, intelligible and easily accessible form, Art. 12 (1) GDPR. This also applies to information obligations regarding automated decision-making [29].

These regulations, however, are not sufficiently clear from a legal point of view [30]. While it shouldn’t be a practical problem to inform about the existence of automated decision-making, it remains unclear what is meant with “meaningful information on the logic involved”. According to the German Federal Court of Justice, an algorithm can be a trade secret, but the data subject has the right to be informed about which personal data is being used to compute a decision [31]. It has to be noted though, that the case was decided in 2013, meaning that the GDPR has not been considered.

Also, the ruling did not consider computational intelligence based systems and the "black box" phenomenon, but an algorithm used for credit scoring. Therefore, the question could be raised as to whether the principles of the ruling could still apply considering the impact of computational intelligence and the GDPR's transparency requirements. For further interpretation of the requirement to provide "meaningful information on the logic involved", recital 63 of the GDPR could be used: It states that the right to access to personal data should not "*adversely affect the rights or freedoms of others, including trade secrets or intellectual property and in particular the copyright protecting the software*." This would, on the one hand, imply that the protection of trade secrets as decided by the German Federal Court of Justice is still a valid argument. On the other hand, it has to be noted that computational intelligence has the potential to hugely affect daily life. Transparency for such decisions is crucial to ensure that people are not stigmatised or discriminated against.

However, comparing automated decision-making to human decision-making – which can, (at present) to an even bigger extent, also affect daily life – it becomes obvious, that human decision-making also lacks transparency. Despite the fact that already from a medical point of view the (intentional or unintentional) true motivation of a human being is not comprehensible, there is no right to such transparency towards human decision-making processes. What we have in different cases is the right to an (ex post) explanation (e.g. court or recruiting decisions). Such obligations, therefore, only refer to the result and not the decision-making process, and do not ensure that a human being – even if he/she was obliged to be transparent – would tell the truth, leaving a risk that a decision would not be transparent but only justifiable. Some analogy could be drawn to the distinction between the "context of discovery" and the "context of justification" as it is discussed in scientific theory [32]. Considering that CI-based systems are increasingly capable of competing with human beings in terms of their ability to interpret information, the question could be raised why computer systems have to be fully transparent, when in fact the principle of full transparency is not applied for interactions between human beings. However, the principle of human dignity ensures that human beings do not have to reveal their inner thoughts, whereas such a right does not exist for machines. While of course a certain degree of transparency is necessary to ensure that a system is not discriminating against people or otherwise violating laws, it will have to be discussed as to how transparent a computer system has to be.

Besides this, information does not necessarily lead to more transparency. Quite the contrary, extensive information often overburdens the person concerned. Therefore art. 12 (1) GDPR states that information must be provided in an intelligible way. Considering the increasing complexity of algorithms and machine learning approaches, this means that it might not be possible to reveal the technical functioning of an algorithm in an intelligible way, but only a simplified description, e.g. on which data is being used and how. This general information most probably is not sufficient to adversely affect trade secrets.

However, with regard to the specific use of deception detection for security purposes such as border control, revealing any information on the functioning of an algorithm, including the categories of personal data, which have been processed, might reveal confidential information about the procedures of security agencies. In these cases, information to be provided might be further restricted and other measures have to be implemented, such as expert groups or ethics commissions.

As shown, the GDPR is only partially capable of addressing the aforementioned issues and it is questionable whether it would be the right place to address these issues. Having a specific regulation on algorithms might be beneficial both for users and end-users of computational intelligence based systems and crucial to guarantee that the rights and freedoms of persons affected are being respected.

3.4 Intelligible information for the data subject

Another legal issue with regard to the information obligations is the requirement to provide intelligible information. This leaves room for various interpretations: Should the information be intelligible for the data controller, for the individual data subject, or rather for an objective, reasonable and informed third party? [33]. In that regard, it needs to be considered that data controllers might have a substantial advantage both in knowledge of their systems and technology in general. While detailed technical information could be on the one hand seen as a maximum level of transparency, an average data subject will most probably not be able to understand such information. Therefore, information should be less detailed than it would be theoretically possible to provide, if this ensures that the data subject can actually understand the information. This is also reflected in art. 12 and recital 58, stating that the data subject shall be addressed using clear and plain language. However, it needs to be ensured that such simplified information is sufficient to enable the data subject to understand the impact of this decision on his fundamental rights, including the right to informational self-determination and appointed expert groups capable to verify more detailed information could be introduced to control the reasoning on a more detailed level.

Last but not least, another challenge is the fact that data subjects can have very different background-knowledge helping them to understand information. People who frequently use ICT services might be more familiar with the functioning of algorithms than people who can barely use a computer. If the GDPR is required to ensure that every *individual* data subject understands the logic involved, data controllers would have no legal certainty as to whether they comply with the legal requirements or not. Therefore, the information provided to describe the logic involved should be intelligible for an objective, reasonable and informed third party persons (“the average user”), while at the same time providing as much information as possible.

3.5 Challenges for the technical community

As outlined above, many issues regarding automated decision-making derive from legal obligations. However, there are certain issues which also require input and solutions from the technical community, in particular:

- How to properly assess the legal situation regarding automated decision-making and how to apply proper safeguards?
- How to explain an algorithm without leaking trade secrets?
- How can algorithms based on computational intelligence be explained?
- Can the information on how an algorithm learns be sufficient to understand it's functioning and decision-making?
- Can self-learning algorithms also explain their decision-making, and could this be updated frequently for every user?

4 Case Study: Profiling Traveller's across Schengen Borders

iBorderCtrl, (Intelligent Portable Control System) is a three year H2020 research and innovation project, funded by the European Union, which is developing novel mobility concepts for land border security. The system will enable authorities to achieve higher throughput at the crossing points through faster processing of passengers within vehicles or pedestrians whilst guaranteeing high security levels through targeting criminal activities such as human trafficking, smuggling and terrorism. In addition, the system will aim to reduce the subjective control and workload of human border agents and to increase the objective control through non-invasive automated technologies. The aim is to ensure that travellers have a speedier border crossing through engaging in a pre-registration step [8]. A full description of the project can be found here: http://www.icross-project.eu/. iBorderCtrl features a unique combination of state-of-the-art biometric tools which will provide risk scores to a Risk Based Assessment Tool (RBAT) that will act as an automated decision-maker on the status of the traveller as they arrive at the border crossing point (Green is proceed, Amber is second line check and Red is recommended refusal – a refusal recommendation would require multiple sources of evidence including document checks, biometric checks etc.). It is important to say that iBorderCtrl is a human in the loop system and therefore provides advice to human border guards who ultimately have the final say. The focus in this chapter is on the profiling of travellers deceptive behaviour in the pre-registration step using a psychological profiling system called ADDS (Automated Deception Detection system) [34] and how such a system when deployed in the field, provides numerous challenges if asked to provide an explainable decision to different stakeholders: the research and development team of ADDS, the Border Guards and their managers and the travellers using the system. The next section provides an overview of ADDS.

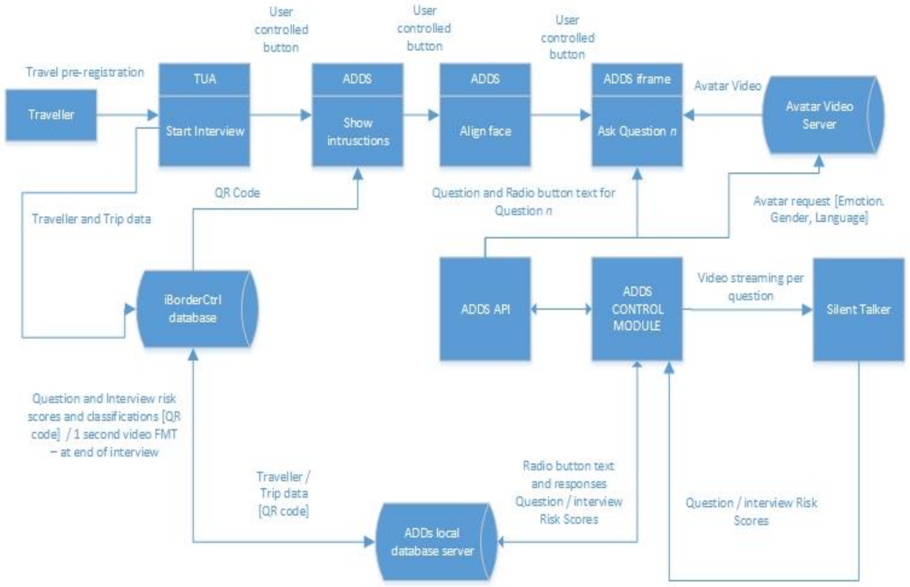

User controlled button
User controlled button
User controlled button
Travel pre-registration
Traveller
TUA
Start Interview
ADDS
Show intrusctions
ADDS
Align face
ADDS iframe
Ask Question *n*
Avatar Video
Avatar Video Server
QR Code
Traveller and Trip data
Question and Radio button text for Question *n*
Avatar request [Emotion. Gender, Language]
iBorderCtrl database
ADDS API
ADDS CONTROL MODULE
Video streaming per question
Silent Talker
Question and Interview risk scores and classifications [QR code] / 1 second video FMT – at end of interview
Traveller / Trip data [QR code]
Radio button text and responses Question / interview Risk Scores
Question / interview Risk Scores
ADDs local database server

Fig. 1. ADDs Component Overview [35]

4.1 Automated Deception Detection System

In the pre-registration phase, after entering the information about their trip each traveller will be required to be interviewed by a Border Guard avatar. Information is exchanged between the iBorderCtrl System and the ADDS system using a unique QR code which is generated for each trip a traveller goes on. An overview of the ADDS components can be found in figure 1. At the start of the interview, the traveller will be provided with instructions on how to align themselves with the camera on their own device. They will then start the interview. The system will present the traveller with one question at a time. After each spoken response, the traveller will be asked to confirm their answer using radio button responses. This acts as a timestamp between questions and responses and is recorded by ADDS. After each question, analysis on deceptive behaviour is undertaken and at the end of the interview, an overall deceptive risk score is calculated. Then this is uploaded to the cloud based iBorderCtrl database where it will make a weighted contribution (based on the ADDS technical readiness level) to the overall risk profile of a traveller.

In the pilot research studies, consenting adults who meet the ethical criteria and agree to take part, will be asked 16 questions, similar to those asked at border crossing points. For example: What will your means of travel into the Schengen Area be? What is the purpose of your trip? The interview will last on average 2.5 minutes subject to network connections and speed. ADDS will be responsible for conducting the interview where the Avatar asks questions, utilising three attitudes (puzzled, neutral and positive) and two avatars, one male and one female which in the first testing phase will be randomly assigned. An example of a female Border Guard avatar (designed by Stremble Ventures [36] is shown in Fig.2. For the purpose of this research and the pilot study, the avatar speaks the question verbally. The traveller then confirms their verbal answer with one extracted from information they have already provided in relation to this question or they can ask the avatar to repeat the question (Fig. 2).

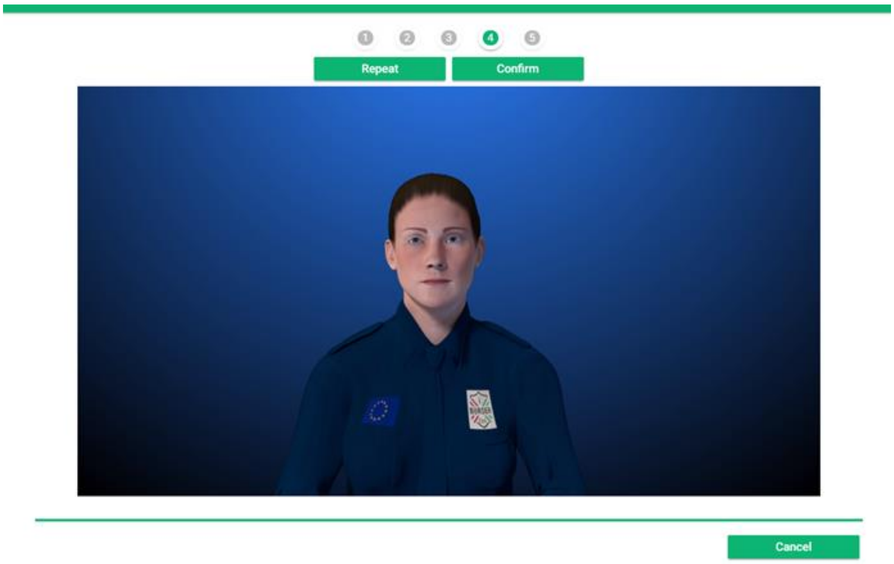
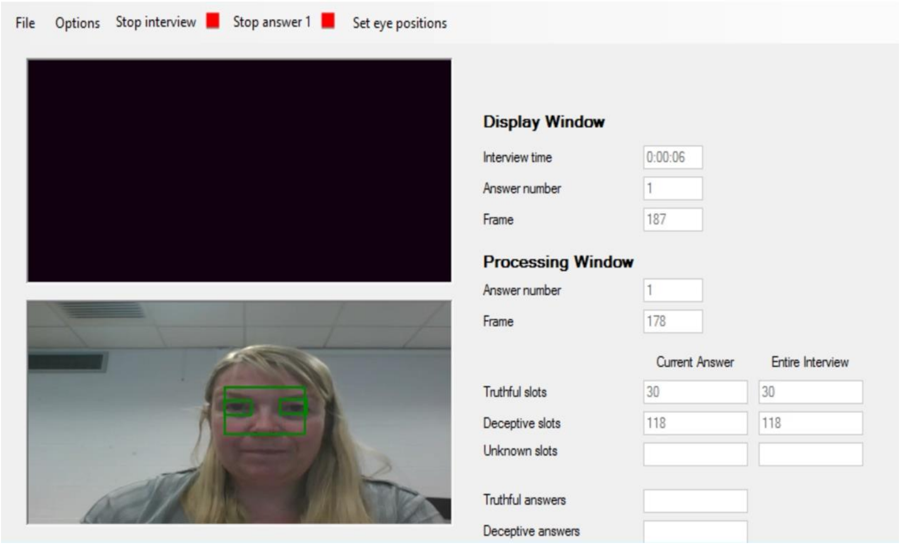


Fig. 2. Female Avatar Border Guard [35].

The non-verbal behaviour of each video question response will be analysed by the Silent Talker system [34] which will output the deceptive risk score for each question. ADDS utilises 38 non-verbal channels which vary in complexity. Each channel is coded to the bipolar measurement range [-1, 1] by the Channel Accumulator [34] and these are ultimately grouped into channel vectors based upon a time slot (i.e. 1 to 3 seconds) before being encoded within the image vector. Figure 3, shows an example of the backend processing carried out by the Silent Talker component of ADDS. The bottom screen, displays the live video stream for a specific interview question, whilst on the right the number of truthful, deceptive and unknown slots are shown. At the conclusion of the interview, questions and interview risk scores and their associated classifications, along with one second’s worth of video frames are uploaded to the iBorderCtrl database to be used by Face Matching Module [37] and RBAT [9]. This prevents attempts to cheat the system by impersonation as the Face Matching Module will check that the traveller who performs the interview is the same as the person on their identity document. Further confirmation of the identity will take place when the traveller reaches the border. In summary, the fundamental level of truthful or deceptive classification is based on a timeslot, a snapshot of non-verbal behaviour during an interview. Timeslots are aggregated over a complete answer to a question to classify the behaviour over the complete answer. Either timeslots or answer classifications may be aggregated over the whole interview to give an overall classification for the interview.



Display Window

Interview time 0:00:06

Answer number 1

Frame 187

Processing Window

Answer number 1

Frame 178

	Current Answer	Entire Interview
Truthful slots	30	30
Deceptive slots	118	118
Unknown slots		
Truthful answers		
Deceptive answers		

Fig. 3. Backend Silent Talker Processing [35].

5 Methodology

Initial experiments were conducted based on a typical airport security scenario of packing a suitcase in order to train, validate and test ADDS for use as a module component in the iBorderCtrl system. An empirical study was conducted using 30 participants whose non-verbal behaviour was recorded whilst engaged in an online interview with a static border guard avatar [34]. The aim of the experiment was to derive models of truthful and deceptive non-verbal behaviour using a configuration of artificial neural networks and traditional decision trees to classify deception and truthfulness.

The hypothesis tested was:

H0: A decision made by an automated deception profiling system cannot be explained using decision tree models

H1: A decision made by an automated deception profiling system can be explained using decision tree models

Each participant first read a participant information sheet and had the ability to ask questions to the researchers, before signing an informed consent form. The participants then took part in a role play exercise which was designed in a similar manner to the suitcase scenario reported in [34]. The activity first involved packing a suitcase with items typically taken on a holiday. Participants were also asked to look at posters typically found at airports showing prohibited items when boarding an aircraft. Each participant was randomly assigned either a truthful or one of four deceptive scenarios designed to cover high and low stakes deception. For example, based on the literature, it was anticipated that a person transporting drugs would generate higher arousal levels than a person being deceptive about packing some illegal agricultural produce. Following the role play, participants were then interviewed by a border guard avatar and were asked 13 questions which are typical of those asked by border guards. This is a different, scenario specific, set of questions from the generic set described for pilot testing in section 4.1. Full details of the experimental methodology are described in [34]. Following data preparation (described in section 2.1), two classification models were developed. One based on the hierarchal ANN model used by Silent Talker and the other using decision tree rule induction which aims to construct a set of rules which will classify objects from knowledge of a set of examples whose classes are previously known. The method is based on recursive partitioning of the sample space and defines classes structurally by using decision trees. The well-known benefit of decision trees is their ability to provide transparency of the decision-making process. For the purpose of this work, Quinlan's C4.5 decision tree algorithm were used [38].

5.1 Data

86584 image vectors were collected from the image data of the 30 participants, where each vector contained the states of each of the 38 non-verbal channels. Ground truth was established for each participant's interview question through knowledge of the scenario that they role-played. I.e. truthful (43051 image vectors) or deceptive (43535 image vectors). Out of the 32 participants, there were 17 deceptive and 15 truthful interviews, 22 males and 10 females with a mix of ethnicities. For the purpose of the ADDS experimental scenario reported in this chapter, the final ANN classifies truthfulness or deceptiveness is based upon an activation level in the range [-1, 1] which was determined from the data set. The deception risk score, *Dq,* of each of the questions was defined as

$$D_q = \frac{\sum_{s=1}^{n} d_s}{n} \qquad (1)$$

Where d_s is the deception score of slot *s* and *n* is the total number of slots for the current question. In order to obtain a classification for each vector, the following thresholds were applied.

```
IF Question_risk ($D_q$) <= x THEN
        Image vector class = truthful (-1)
```

```
ELSE IF Question_risk ($D_q$) >= y THEN
        Image vector class = deceptive (+1)
ELSE
        Indicates not classified
END IF
```

Where x = -0.05 and y = +0.05. Thus if a question risk score was within this range, a classification could not be allocated. This range is empirically defined, determined through previous work and only used for the initial experimental scenarios described in [34].

5.2 Results and Analysis

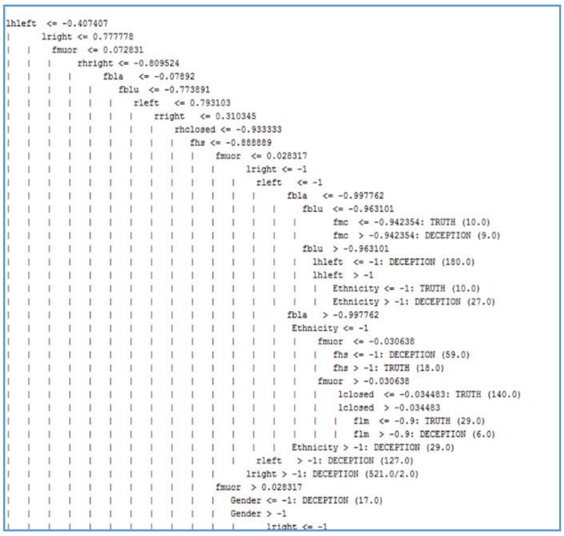
In [34], two methods for training, validation and testing were reported: *n-fold* cross validation and leave a pair out. The latter being a more appropriate measurement of accuracy for unseen participants, which is required when a system such as ADDS is deployed in the field. However, for the purposes of this chapter, in the context of comparing models in terms of classification accuracy and their ability to produce an explainable decision, cross validation is used, as initial work showed there was little difference in induced decision tree size. With no ANN or C4.5 optimisation, the best tree from performing *10-fold* cross validation contained 1072 rules. Table 1 shows the overall classification accuracy of 10-fold cross validation for both the ANN (ADDS-ANN) and C4.5 (ADDS-DT). The additional results of the probabilistic classifier Naïve Bayes are also shown for comparative purposes.

Table 1. 10-fold cross validation results

Method	%Train AVG	%Test AVG	%Class-Accuracy
ADDS-ANN	97.03	96.66	96.80
ADDS-DT	98.90	98.80	98.80
ADDS-Naïve Bayes	80.12	70.12	75.12

5.2.1 Are Decisions Explainable?

Fig.4. shows a snapshot of the best decision tree which contained 1072 rules and had a tree size of 2143 nodes.



```
lhleft  <= -0.407407
|   lright <= 0.777778
|   |   fmuor  <= 0.072831
|   |   |   rhright <= -0.809524
|   |   |   |   fbla   <= -0.07892
|   |   |   |   |   fblu  <= -0.773891
|   |   |   |   |   |   rleft   <= 0.793103
|   |   |   |   |   |   |   rright   <= 0.310345
|   |   |   |   |   |   |   |   rhclosed <= -0.933333
|   |   |   |   |   |   |   |   |   fhs <= -0.888889
|   |   |   |   |   |   |   |   |   |   fmuor  <= 0.028317
|   |   |   |   |   |   |   |   |   |   |   lright <= -1
|   |   |   |   |   |   |   |   |   |   |   |   rleft   <= -1
|   |   |   |   |   |   |   |   |   |   |   |   |   fbla   <= -0.997762
|   |   |   |   |   |   |   |   |   |   |   |   |   |   fblu  <= -0.963101
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   fmc  <= -0.942354: TRUTH (10.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   fmc  > -0.942354: DECEPTION (9.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   fblu  > -0.963101
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   lhleft  <= -1: DECEPTION (180.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   lhleft  > -1
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Ethnicity <= -1: TRUTH (10.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Ethnicity > -1: DECEPTION (27.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   fbla   > -0.997762
|   |   |   |   |   |   |   |   |   |   |   |   |   |   Ethnicity <= -1
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   fmuor  <= -0.030638
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   fhs <= -1: DECEPTION (59.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   fhs > -1: TRUTH (18.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   fmuor  > -0.030638
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   lclosed  <= -0.034483: TRUTH (140.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   lclosed  > -0.034483
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   flm  <= -0.9: TRUTH (29.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   flm  > -0.9: DECEPTION (6.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   Ethnicity > -1: DECEPTION (29.0)
|   |   |   |   |   |   |   |   |   |   |   |   rleft   > -1: DECEPTION (127.0)
|   |   |   |   |   |   |   |   |   |   |   lright > -1: DECEPTION (521.0/2.0)
|   |   |   |   |   |   |   |   |   |   fmuor  > 0.028317
|   |   |   |   |   |   |   |   |   |   |   Gender <= -1: DECEPTION (17.0)
|   |   |   |   |   |   |   |   |   |   |   Gender > -1
|   |   |   |   |   |   |   |   |   |   |   |   lright <= -1
```

Fig. 4. Rule Snapshot [35].

The rules induced from the dataset represent patterns of non-verbal behaviour for specified channels which, when combined, allow the classification of deception verses truth for a given risk score. One rule from this tree which gives a classification of deception can be extracted as follows:

> *IF lhleft < -0.407407 AND lright <= 0.777778 AND fmuor <=0.072831 AND rhright <= 0.310345 AND rhclosed <=-0.93333 AND fhs <= -0.888889 AND fmour <=0.028317 and lright <=-1 and rleft <=-1 and fbla <=-0.997762 and fblu <=-0.963101 and fmc > -0.942354 THEN CLASS DECEPTION.*

Note that this is a relatively simple rule as it deals with summary statistics for each channel. Analysing the rule, one familiar with the underpinning theory of Silent Talker will see information on four non-verbal channels associated with the eyes: left eye looking left (lleft), left eye looking right (lright), right eye half closed (rhclosed), right eye looking left (rleft) and 5 channels containing information about the state of the face including the horizontal movement of the face (fhs) face angular movement up-on-right (fmuor) and the degree of blushing/blanching (fblu). Face channels track the face movement along the X-axis and Y-axis using the coordinates and dimensions of the face found by the Face Object Locator ANN [7]. Likewise, the state of each eye channel is determined from a Pattern Detector ANN [34] observing the left/right eye image and/or from the application of logical decision(s). The values for each channel are determined empirically by the pattern detector ANNS and the channel encoder ANNS in the bipolar range [1 and -1].

In this application, the rules are complex and look at combinations of fine grained non-verbal behaviour i.e. movement of facial features. Due to this complexity, individual rules are difficult for an average human to comprehend. They could not for example be replicated by a human. As the problem is complex, the tree is large –

previous work [39] suggests pruning may lead up to a 25% reduction in rules. A sacrifice in classification accuracy occurs but still the quantity of rules is large and difficult to comprehend. But is this problem scenario based? If automated profiling were applied to a simpler more typical problem, such as a bank loan or mortgage application then perhaps the learned rules could be understood by all stakeholders – the expert, the member of the public and the bank manager. Consider for example, a small dataset containing 434 instances for applications for personal loans. 238 instances are reject samples and 196 accept. The dataset contains just 14 attributes. Using C4.5 and *10-fold* cross validation, a classification accuracy of 74.8% is achieved and the best tree contains 27 rules. A sample rule is shown below:

> *IF TimeAtBank(years) <=2 AND TimeEmployed(years) < 1 AND ResidentStatus = "Rented" THEN Outcome = REJECT LOAN.*

A person, profiled by this system, could have the decision explained to them using this rule by a staff member at the bank i.e. they had not been a customer at the bank for long enough and had not been in employment for over a year and they currently lived in rented accommodation. What the staff could not do is qualify the accuracy of the model used to train the system, explain and show the statistical evidence behind the decision, nor guarantee that there was any bias in the training data that led to the model. Therefore, neither the hypothesis H0 nor H1 can be universally accepted as explainability is determined by problem representation and complexity.

5.2 The Future of Explainable Decision Making

Section 5.1 provides evidence that, at least at the most detailed technical level, some Computational Intelligence (CI) decision making will be completely inexplicable even to a reasonably well-informed and educated public. Still from a legal point of view, there is no distinction in the GDPR with regard to the information obligations as to how complex a decision is. However, this should not provide reasons for abandoning the use of CI or the duty of explainability. We propose a progressive approach to explaining decisions and challenging them, based on a hierarchy of capability. Understanding information at a particular level of complexity will result in a particular capability to question and object – and data subjects should have a right to question, challenge and receive answers appropriate to their level of attainment. The proposed Hierarchy of Explainability and Empowerment is shown in Fig.5.

The image part with relationship ID rId13 was not found in the file.

Fig. 5. O'Shea's Hierarchy of Explainability and Empowerment

There is a clear disjunction in the hierarchy. One might question whether the top three levels can be intelligible at all, whereas reaching the consent level should be considered as the minimum requirement for a CI system making decisions automatically. At present, the best approach to ensure transparency seems to be to maximise readability of text explanations and minimise the mathematical complexity of formulae in explanations to data subjects. Recital 63, also covering trade secrets or intellectual property (see section 3.3) may prevent a full technical disclosure pertinent to the upper levels of the hierarchy. However, taking into account the abovementioned arguments to what extent the right to information shall be interpreted, an answer could be the following: From a technical point of view, there is the possibility of creating a simpler abstraction, although in the area of “know-how” (such as data cleansing) even procedures that appear quite simple may need to be protected for commercial reasons. On the other hand, anything which is patented is automatically fully-disclosed (see section 5.2.5). Each level will now be explained.

5.2.1 The Context Level

This level considers the Computational Intelligence systems as component of a multiplex (a collection of systems), which itself should be contestable at a societal level. The multiplex should be understandable and the overall process of decision making should be challengeable by data subjects as citizens, through (usually elected) representatives with a duty, time and resources to be better informed than the general public and legislate on their behalf for example, see [40]. In one sense, the actions at this level are the drivers or enablers for the Consent level. As the multiplex level may contain multiple black boxes, it is legitimate for the public to express themselves in terms of subjective feelings and decisions at this level. Decision-making will largely rely on verbal reasoning and debate. Because the particular CI component is part of an overall multiplex which may contain other CI and Information Systems, the data subject should be entitled to know which components were influential in making the decision and how their contributions were combined. In particular, the subjects should be informed whether the CI system was influential in the decision taken about them and if so, which of the other systems in the multiplex confirmed or contradicted the CI system. At this level, it should be established that the data subject should be allowed to contest the decision without having to make a technical argument.

5.2.2 The Consent Level

This level concerns the ability to understand the personal consequences of using the CI system, to give consent or opt out and to object to an automated decision. As this is crucial to allow for a lawful processing on common legal bases such as consent (art. 6 (1) lit. a. GDPR) or legitimate interests (art. 6 (1) lit. f. GDPR), this level should be seen as the minimum requirement regarding information provided to the data subject. It should provide information about the basic quantitative measures used by the system. For example, the percentage accuracy in making decisions with contextual information about humans making the same decisions.

An example invitation to do the pre-travel interview in the border control context might be:

“I am inviting you to do a pre-travel interview. This could make your border crossing faster.
A Computational Intelligence system will watch your interview and calculate the risk that you were lying.
The system calculates other risks, for example from a machine reading your passport.
The lie detector is not 100% accurate, we know this and allow for it when we calculate risks.
No lie detection method, including human experts, is 100% accurate.
The current accuracy of the system is 75%.[1]
If the combined risks are high, you will have an interview with a border guard to clear things up.
Would you like to take the pre-travel interview?”

This passage has Flesch Readability Ease score of 60.9 and a Flesch-Kincaid Grade Level of 7.8, and is described as “Plain English. Easily understood by 13- to 15-year-old students.” (USA)

An example notification of the right to contest in the border control domain might be a statement such as:

“Your answers in the pre-travel interview showed there was a risk that you weren’t telling the truth.

[1] We expect to achieve 85% accuracy if such a system were deployed.

The passport scanner had difficulty reading your passport.
The border control system thinks there may be a risk if allows you to cross the border.
Machines can make mistakes. The current accuracy of the deception detection system is 75%.
This is your chance to clear things up by speaking to a human border guard."

This passage has Flesch Readability Ease score of 74.8 and a Flesch-Kincaid Grade Level of 5.8, and is described as "Fairly easy to read." In the border control example, the text should be accompanied by a contingency table (not using technical jargon such as "false alarm"), for example,

"These are the results of our tests of the deception detector component. You can see how well it performs with truthful people and deceptive people that it has never seen before. Errors are shown in red. These errors show you more about how the CI system could have made a mistake in interviewing you and may help you decide whether or not you wish to contest the decision."

Truthful cases	Deceptive cases	
75.55	*24.55*	Classified as truthful
26.34	*73.66*	Classified as deceptive

A contingency table helps the data subject to understand whether the CI classifier is biased towards one particular outcome or not. It is worth noting, that the results shown in the contingency table are from preliminary experiments conducted in 2017 on 30 people and published in [34]. Understanding the contingency table requires numeracy skills of understanding and reasoning with percentages and proportions, corresponding to upper key stage 2 of the UK national curriculum, year 6 (aged 10 and above).
This level cannot explain the logic of a CI system as a detailed algorithm, but it does explain how the system reasoned about its calculations and relates this reasoning to the inputs, for data subjects with a general education. While this would probably allow the data subject to better decide on whether he/she would like to ask for human intervention, it would probably not provide meaningful information on the logic involved in the algorithm, as required by the GDPR.

5.2.3 The Comprehend Level

This level concerns the ability to integrate understanding of various aspects of the CI system to envisage their consequences for decision making. It involves some disclosure of the logic involved in the automated decision-making, but involves a tradeoff between transparency & accessibility on one hand and providing meaningful information on the other.
This requires drilling down into the black box system to reveal it at unit level (as far as possible). It supports understanding the relationships between components (units) working together in the CI system. Data subjects should be able to compose a more

informed question or objection about the correct performance of a component in the overall system.
Numeracy skills will become more important at the comprehend level, however it should be possible to provide meaningful quantitative descriptions within the scope of upper key stage 4 of the UK national curriculum (years 10 &11, aged 14 – 16),
Considering the pre-travel interview for border control, the data subject may be informed of the roles of components, their inputs and outputs, and how the data flows between them to make an overall classification, provided there is no conflict with Recital 63 IP Protection. Concepts such as signed numbers, thresholding and inequality relations characterize reasoning about these components.
An *example notification* of the right to contest in the border control domain might be a passage such as:

"The Computational Intelligence system breaks each question up into a series of short video clips (slots).
Each video clip is processed by the system and given a risk score between -1 (minus one) and +1 (plus one).
-1 shows a strong belief that you were truthful.
+1 shows a strong belief that you were deceptive.
Numbers in between show different strengths of belief.
The CI system then checked if the risk score was less than -0.05. If it was, it classified your video clip as truthful.
Otherwise, it tested if your answer was greater than +0.05. If it was, it classified your slot as deceptive.
For risks between -0.05 (very weak truthful) and +0.05 (very weak deceptive) it decided it was not confident enough to classify the slot.
The lie detector is not 100% accurate and some answers will be classified incorrectly, but it ignores slots that contain weak indicators and it calculates the score for the complete answer by comparing the number of deceptive video clips to the total number of video clips for an answer.

The classifications for your answers were:

1. *What is your surname? – Deceptive*

(etc.)

Machines can make mistakes. You may have good reasons and evidence to challenge these classifications.
This is your chance to clear things up by speaking to a human border guard"

This passage has Flesch Readability Ease score of 63.8 and a Flesch-Kincaid Grade Level of 7.6, and is described as "Plain English. Easily understood by 13- to 15-year-old students." The numeracy requirements to understand and challenge at this level include concepts such as signed numbers, thresholding and inequality relations (section 5.1). This correspond to upper key stage 4 of the United Kingdom national curriculum, years 10 &11 (aged 14 – 16).

It should be noted that, in the examples in this section, the interview answers given by the data subject have been classified (to some degree) as deceptive. Formally, this does not preclude a data subject from challenging a favourable (i.e. truthful) classification by the system.

5.2.4 The Confidence Level

This level discloses information about processes within the CI system. At this level, there is more likely to be conflict between protecting the Intellectual Property in the CI system and informing the data subject. The intention is to demonstrate that their data is represented accurately, that appropriate procedures are applied and that consequently the risk of a profiling error has been minimised. Data representation is, superficially, the simplest element. The data used to make the decision can be disclosed for the subject to check allowing them to be sure that they have not been misidentified or to provide evidence to correct errors of fact in records. Some data may not be disclosable for reasons of security – however, this is not a particular issue for a CI system since the same would apply if the data were processed by human border guards (for example, information in a law enforcement database obtained through intelligence gathering). However, the data belonging to a data subject undergoes mathematical transformations when processed by a CI system, some may be explainable at this level, others may only be explainable at the contest level.

Some performance figures have already been disclosed at the lower levels of the hierarchy. Understanding can be enhanced at the confidence level with statistical analysis – not only did the CI component perform with a certain accuracy during testing, but this is the level of confidence that the figure quoted did not just occur by chance.

A further disclosure is the demographic composition of data sets used to train and test the system currently deployed. This will allow data subjects to challenge the result of the test based on lack of representation of their gender, ethnicity or other possible confounding factors. Data subjects viewing the system at this level will need to understand concepts such as statistical significance, α and p values, population samples, frequency distributions etc. to make effective use of the information provided. The mathematical skills required correspond to AQA A-level statistics in the UK (aged 16-18).

An *example notification* of the right to contest in the border control domain might be a passage such as:

> *"The Computational Intelligence system analysed the video of your answers to questions by locating non-verbal features (mainly facial features), detecting their states and combining these, over time, to feed a final classifier.*
> *In this system all of the tasks were performed by Artificial Neural Networks.*
> *ANNs are created using sets of training and testing data, from which they learn solutions that apply to cases they have never seen.*
> *Therefore, they require representative data sets.*
> *The data sets used in iBorderCtrl were composed of (percentages for gender distribution, percentages for ethnic distribution).*

> *The ANNS trained for locating and detecting states of features achieved at least x% accuracy*
> *The ANN for finally classifying deception achieved 75% accuracy*
> *The p-values the ANNs detecting and locating were all less than the α value of 0.05, the normal acceptable level for scientific tests. Some were less than α value of 0.01 which is the stronger level for scientific testing.*
> ***The p-value for the ANN finally classifying deception was n.nn which is less than** the α value of p.pp etc.*
> *Machines can make mistakes. You may have good reasons and evidence to challenge these classifications, this is your chance to clear things up by speaking to a human border guard."*

This passage has Flesch Readability Ease score of 49.9 and a Flesch-Kincaid Grade Level of 10.8, and is described as "Difficult to read" and "Sophomore" level in the education system (USA). The numeracy requirements to understand and challenge at this level include concepts such as statistical significance, α and p values, population samples, frequency distributions etc. to make effective use of the information provided. The mathematical skills required correspond to AQA A-level statistics in the UK (aged 16-18).

5.2.5 The Contest Level

The Contest level discloses information at the atomic level, allowing a sufficiently skilled data subject with state-of-the-art understanding of the CI system to contest its decisions in detail. This level of disclosure assumes that there is no conflict with Recital 63. It should be noted that Recital 63 will not always be an obstruction. Gaining patent protection for commercial use of such a system requires sufficient information to be disclosed for "one sufficiently skilled in the art" to reproduce it.
At this level, the data subject should be able to express a fully-informed point of view backed by evidence at the level of algorithm functioning. Data subjects will need advanced understanding of CI and the domain, for example topics such as calculus, entropy and transfer functions. This would typically require education to honours degree level (including specialist CI units of study) or PhD level. Data subjects with this level of skill and education would be informed through peer-reviewed scientific publications and patents. Theoretically, this level would enable the data subjects to track their data through the system, replication of the mathematical operations, checking their results and tracing the steps to the final classification. This would have little or no value to the typical data subject due to the infeasibility in terms of time for the human to replicate the computer calculations on their data and the opaqueness of the classification (we know what happened, but we don't know why it happened).The real value of this level of disclosure is in the protection offered by expert scrutiny and validation of the system by independent scientists.

5.2.6 Concluding meaningful vs. intelligible information

The various levels as described above outline the differences in the level of detail which could be applied when informing a data subject on the logic involved when processing personal data, including automated decision-making. As described in section 3.4, this information, however, also must be intelligible. While transparency is crucial to enable the data subject to conclude informed decisions, a high level of detail, which the data subject might not be able to understand, does in fact not increase transparency, but rather is contraindicative. Similar to legislation on consumer protection, information obligations should focus on being short and straight to the point, while including all relevant information.

In general, the consent level can be seen as minimum requirement for any information obligation deriving from a data processing operation. In most use-cases relating to GDPR, the data subject needs to be enabled to make informed decisions. However, for certain use-cases such as for border checks or other security related matters, revealing information on the logic involved might need to be limited quite exhaustively, even to an extent below the consent level. While in that case the context level might be still sufficient to generally inform the data subject on the fact that automated decision-making is being used, additional safeguards would be needed. Such a data processing operation would in any case require a statutory legal basis, but also additional safeguards such as auditing algorithms and oversight mechanisms to ensure that, even though the data subject's rights are restricted, the risk of violations of fundamental rights is minimised.

Regarding uses-cases without such restrictions, it needs to be monitored how the understanding and education of average users will develop in the future. While for the most users intelligible information nowadays might be covered by the consent or comprehend level, this might change over time, requiring to also include information from the comprehend, confidence and contest level. Consequently, the concept of information obligations has to be understood as a flexible model, allowing for adjustments and refinements whenever required.

Depending on the impact and relevance on the data subject, certain use-cases might also require an additional monitoring of algorithms used for automated decision-making through specialised entities such as NGOs, expert groups or oversight authorities. While such experts could also understand the top-levels of the proposed model, the risk of infringements of the data controller's interests could be minimised, in particular regarding IP protection, while at the same time ensuring that an algorithm does not violate the data subject's fundamental rights without him/her being able to actually realise this.

6 Conclusions

This chapter has used a case study of adaptive psychological profiling to examine the challenges of how to produce explainable decisions of CI models to all stakeholders. There are many challenges for both the computational intelligence and the legal communities. Therefore, finding solutions that reflect technical realities while at the same time providing sufficient privacy safeguards is crucial. This, however, requires a close collaboration of both the legal and technical community, which currently happens very rarely. A closer collaboration between both communities would allow better guidance, such as common guidelines for software developers, standardized frameworks which comply with the GDPR by default, and many more [41]. Therefore, receiving answers to the questions raised in section 3.5 would be an important first step to further deepen the common understanding of technical and legal challenges relating to the GDPR and to foster a debate on the proper interpretation of the GDPR among the legal community, as well as information on how the technical community could be supported in their efforts to comply with legal requirements. Finally, the future of explainable decision making is expressed as a new proposed Hierarchy of Explainability and Empowerment which allows information and decision making complexity to be explained at different levels depending on the abilities of a person. Examples are given of the hierarchy for responding to different needs of explainability. Such a hierarchy can be seen as the conceptual starting point in addressing explainability in CI systems.

Acknowledgements

The iBorderCtrl project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 700626. Silent Talker research has been supported by Silent Talker Ltd.

References

1. Bonn, S. 2017. Criminal Profiling: The Original Mind Hunters, Profiling killers dates back to Jack the Rip-per, In psychology Today, [online] https://www.psychologytoday.com/us/blog/wicked-deeds/201712/criminal-profiling-the-original-mind-hunters [Accessed 27/7/2018]
2. Silent Talker Ltd. 2019 [online] https://www.silent-talker.com/. [Accessed 13/01/2019]
3. Bryan Casey, B., Farhangi, A & Vogl, R, 2018. Rethinking explainable machines: the GDPR’s “right to explanation” debate and the rise of algorithmic audits in enterprise, [online] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3143325 [Accessed 21/01/2019].
4. GDPR Portal, 2018, [online]. Available at: https://www.eugdpr.org/ [Accessed 27/07/2018].
5. Bandar, Z. Rothwell, J. McLean, D. O’Shea, D, 2004. Analysis of the behaviour of a subject, Publication date 2004/5/3, Patent office US, Application number 10475922
6. Holmes, M. Latham, A. Crockett, K, O’Shea, J. 2017. Near real-time comprehension classification with artificial neural networks: decoding e-Learner non-verbal behaviour, In IEEE Transactions on Learning Technologies, Issue: 99, DOI: 10.1109/TLT.2017.2754497.
7. Buckingham, F. Crockett, K. Bandar, Z. O'Shea, J. MacQueen, K. Chen, M. 2012. Measuring Human Comprehension from Nonverbal Behaviour using Artificial Neural Networks, Proceedings, In Proceedings of IEEE World Congress on Computational Intelligence Australia, pp368-375, DOI: 10.1109/IJCNN.2012.6252414
8. Crockett, KA and O'Shea, J and Szekely, Z and Malamou, A and Boultadakis, G and Zoltan, S, 2017. Do Europe's borders need multi-faceted biometric protection? In Biometric Technology Today, (7). pp. 5-8. ISSN 0969-4765
9. iBorderCtrl Intelligent Portable Control System [online], Available at http://www.iborderctrl.eu/ [Accessed 07/01/2019].

10. Ganis, G., Kosslyn, S.M., Stose, S., Thompson, W.L. and Yurgelun-Todd, D.A., 2003. Neural correlates of different types of deception: an fMRI investigation. Cerebral cortex, 13(8), pp.830-836.
11. Feldman, R.S., Forrest, J.A. and Happ, B.R., 2002. Self-presentation and verbal deception: Do self-presenters lie more? In Basic and Applied Social Psychology, 24(2), pp.163-170.
12. Trovillo, P. 1939. History of Lie Detection, In Journal of Crim. L. & Criminology, 848.
13. International League of Polygraph Examiners. 2016.), Polygraph/Lie Detector FAQs. [Online] http://www.theilpe.com/faq_eng.html. [Accessed 28/7/2018]
14. Miller, G. The magical number seven, plus or minus two: Some limits on our capacity for processing information 63 (2), 81-97, (1956).
15. Caso, L., Gnisci, A., Vrij, A. and Mann, S., 2005. Processes underlying deception: an empirical analysis of truth and lies when manipulating the stakes. In Journal of Investigative Psychology and Offender Profiling, 2(3), pp.195-202.
16. MacLaren, V. 2001. A quantitative review of the Guilty Knowledge Test. In Journal of Applied Psychology, Vol 86(4), Aug 2001, pp. 674-683
17. Sen, T., Hasan, M.K., Tran, M., Yang, Y. and Hoque, M.E., 2018, May. Say CHEESE: Common Human Emotional Expression Set Encoder and its Application to Analyze Deceptive Communication. In Automatic Face & Gesture Recognition (FG 2018), 2018 13th IEEE International Conference on (pp. 357-364).
18. Rothwell, J. Bandar, Z. O'Shea, J. McLean, and D. 2006, Silent talker: a new computer-based system for the analysis of facial cues to deception, Journal of Applied Cognitive Psychology, Volume 20, Issue 6, pages 757–777.
19. Von Lewinski, 2018. In BeckOK DatenschutzR DS-GVO Art. 22, Fn. 28. [online] https://beck-online.beck.de/?vpath=bibdata%2fkomm%2fBeckOKDatenS_26%2fEWG_DSGVO%2font%2fBECKOKDATENS%2eEWG_DSGVO%2eA22%2eglB%2eglIV%2egl1%2ehtm. [Accessed 27/07/2018]
20. Art. 29 Data Protection Working Party, 2018. WP251rev.01, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, p. 21 f. [online] http://ec.europa.eu/newsroom/article29/document.cfm?doc_id=49826 [Accessed 21/01/2019].
21. Von Lewinski, 2018b. In BeckOK DatenschutzR DS-GVO Art. 22, Fn. 2. [online] https://beck-online.beck.de/?vpath=bibdata%2fkomm%2fBeckOKDatenS_26%2fEWG_DSGVO%2font%2fBECKOKDATENS%2eEWG_DSGVO%2eA22%2eglA%2ehtm. [Accessed 27/07/2018]
22. Martini,2018b. In Pal/Pauly DS-GVO 2nd Edition 2018, Art. 22, Fn. 20a. [Accessed 27/07/2018]
23. Fundamental Rights Agency, 2018a. Preventing unlawful profiling today and in the future: a guide, p. 12. [Online] https://fra.europa.eu/en/publication/2018/prevent-unlawful-profiling. [Accessed 17/01/2019]
24. Fundamental Rights Agency, 2018a. Preventing unlawful profiling today and in the future: a guide, p. 60. [Online] https://fra.europa.eu/en/publication/2018/prevent-unlawful-profiling. [Accessed 17/01/2019]
25. Borges, G., 2018. Rechtliche Rahmenbedingungen für autonome Systeme, NJW 2018, 977 (979).
26. Bundesregierung (German federal government), 2018. Klarheit und Kontrolle bei Algorithmen, [online] https://www.bundesregierung.de/Content/DE/Artikel/2017/07/2017-03-07-maas-algorithmen.html [Accessed 26/11/2018].

27. European Parliament. 2017. European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)), [online] http://www.europarl.europa.eu/sides/getDoc.do?type=TA&language=EN&reference= P8TA-2017-0051 [Accessed 26/11/2018].
28. The European Commission's high-level Expert Group on Artificial Intelligence, 2018. Draft Ethics Guidelines for Trustworthy AI. [online] https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=56433 [Accessed 21/01/2019].
29. Martini, 2018c. In Paal/Pauly DS-GVO 2nd Edition 2018, Art. 22, Fn. 41a.
30. Bräutigam/Schmidt-Wudy, (2015) Das geplante Auskunfts- und Herausgaberecht des Betroffenen nach Art. 15 der EU-Datenschutzgrundverordnung, CR 2015, 56 (62).
31. German Federal Court of Justice, 2014. VI ZR 156/13. [online] http://juris.bundesgerichtshof.de/cgi-bin/rechtsprechung/document.py?Gericht=bgh&Art=en&nr=66910&pos=0&anz=1 [Accessed 21/01/2019].
32. Schickore, J. & Steinle, F., 2006. Revisiting Discovery and Justification. Historical and Philo-sophical Perspectives on the Context Distinction, DOI: 10.1007/1-4020-4251-5.
33. Schmidt-Wudy, 2018b. In BeckOK DatenschutzR DS-GVO Art. 15, Fn. 78. [online] https://beck-online.beck.de/Dokument?vpath=bibdata%2Fkomm%2Fbeckokdatens_24%2Fewg_dsgvo%2Fcont%2Fbeckokdatens.ewg_dsgvo.a28.htm. [Accessed 27/07/2018]
34. O'Shea, J. Crockett, K. Khan, Kindynis, P. Antoniades, A. Intelligent Deception Detection through Machine Based Interviewing, In Proceedings of IEEE International Joint conference on Artificial Neural Networks (IJCNN), DOI: 10.1109/IJCNN.2018.8489392
35. Crockett, K. Stoklas, J. O'Shea, J. Krügel, T. Khan, W. (2018), Adapted Psychological Profiling Verses the Right to an Explainable Decision, 10th International Joint Conference on Computational Intelligence, Seville, Spain, ISBN: 978-989-758-327-8
36. Stremble Ventures. (2019) [online] http://stremble.com/. [Accessed 28/11/2019]
37. Rodiguez, L. Hupoint, I. Teno, C. 2018. Facial Recognition Application For Border Control, In Proceedings of IEEE International Joint conference on Artificial Neural Networks (IJCNN), DOI: 10.1109/IJCNN.2018.8489113
38. Quinlan, R. 1994. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers. ISBN 1-55860-238-0.
39. O'Shea, J. Crockett, K. Khan, W. (2018) A hybrid model combining neural networks and decision tree for comprehension detection, In Proceedings of IEEE International Joint conference on Artificial Neural Networks (IJCNN), DOI: 10.1109/IJCNN.2018.8489621.
40. O'Shea, J. "Dr James O'Shea - written evidence" (to the UK House of Lords Artificial Intelligence Select committee), AIC0226, published 26 October 2017, http://www.parliament.uk/business/committees/committees-a-z/lords-select/ai-committee/publications/
41. Crockett, K. Goltz, S. Garratt, M. 2018. GDPR Impact on Computational Intelligence Research, In Proceedings of IEEE International Joint conference on Artificial Neural Networks (IJCNN), DOI: 10.1109/IJCNN.2018.8489614