BAE Systems plc — Written evidence (NTL0056)

Introduction

This document is BAE Systems' response to the House of Lords Justice and Home Affairs Select Committee's inquiry into "new technologies and the application of the law". In this this document, we discuss our perspective on artificial intelligence (AI) in policing and justice as well as answer the specific questions posed to us.

At BAE Systems, we provide some of the world's most advanced, technology-led defence, aerospace and security solutions. We employ a skilled workforce of 89,600 people in more than 40 countries. Working with customers and local partners, we develop, engineer, manufacture, and support products and systems to deliver military capability, protect national security and people, and keep critical information and infrastructure secure.

This submission has been prepared by the BAE Systems Applied Intelligence business. BAE Systems Applied Intelligence delivers cyber security and intelligence solutions which help our clients protect and enhance their critical assets. Building on our strong heritage of security and defence, our solutions and services help nations, governments and businesses around the world defend themselves against cybercrime, reduce their risk in the connected world, comply with regulation and transform their operations.

BAE Systems has experience with AI solutions across defence, security, finance, and policing. In policing and justice we help support the delivery of operational benefits from advanced data analytics and AI by acting as trusted advisors and delivery partners to police forces and the Home Office. We have a strong interest in AI R&D and innovation, as well as the ethical use of the technology. We believe our experience makes our response relevant to your Inquiry.

We are aware of the challenges and opportunities of police AI initiatives, such as those at Durham Constabulary (Harm and Risk Assessment Tool) and West Midlands Police (National Data Analytics Solution, sponsored by the Home Office). We have piloted advanced data analytics technology with Gloucestershire Constabulary. We recently (December 2020) authored a white paper on AI-led Policing [1], where we discussed the ethical and governance issues facing the use of AI in policing and justice.

We understand the enormous potential and the ongoing risks of AI in policing and justice, and recognise the wide variety of potential use cases – from those that are low risk, such as data triage, through to higher risk use ones such as facial recognition, which could have significant societal impact.

Principles

To frame our response we outline some principles that are relevant to the use of AI in policing and justice, as well as wider public sector use. We will address the specific questions posed by the Committee later in this document.

Rooted in ethics

Ethics should be a primary and ongoing consideration in the design and deployment of any AI solution. It is very important to regularly assess whether the perceived benefits justify the potential risks that can arise – including in particular whether the use of AI is necessary, proportional and whether outcomes will be fair and unbiased to all citizens, accountable and auditable.

We believe an ethical framework should be used for this assessment. The Alan Turing Institute (ATI) framework [2] is an excellent guide although it may need specific tailoring for policing and justice. We fully support the FAST principles of fairness, accountability, sustainability and transparency developed within the ATI framework. The ethical framework should drive an appropriate assessment of risks, as well as the standard of assurance and design documentation required.

We also recommend the use of independent ethics committees to govern decisions on ethical use: both at concept stage and throughout the solution lifecycle. We have seen this successfully implemented at West Midlands Police.

At West Midlands Police (WMP), an independent and diverse Ethics Committee was set up to advise the Police and Crime Commissioner and the Chief Constable. The Committee's remit includes ethical governance of AI and data analytics capabilities, and it publishes its minutes to maximise transparency and generate public trust. The Committee has real impact: it actively reviews AI and analytics projects and has ensured that project direction and approach is fully aligned to ethical considerations. Other forces, such as Sussex and Surrey, have since created ethics committees which follow the WMP model.

Augmenting human decision-making

In some scenarios AI technology supports human decision making, and in others it drives an automated assessment or outcome. With current technology maturity, we believe that for higher risk applications that could significantly impact the public having a human-in-the-loop is an important general principle: in this case the technology is augmenting rather than replacing user judgement. User (and their organisation's) judgement is needed to validate outcomes, mitigate errors, and be accountable for decisions in order to support public trust. As technology, ethical and quality assurance maturity evolves this balance will shift so that there will be increased reliance on, and trust in, completely automated solutions.

We recognise there is a potential problem with decision bias in the human interpretation of results, where users may blindly trust the AI output, or alternatively ignore it based on their own preconceptions. We recommend this should be addressed through training, skills enhancement, transparency or explainability of outputs (wherever possible) and a good user experience.

Standardisation

Standards and frameworks (or guidelines) will be very important for the implementation of AI for policing and justice. AI implementation is in its infancy, there is a diversity of expertise and a distributed user and procurement base across the various police forces, and there are lots of potential use cases – many of which are inherently high risk. Standards and frameworks provide a baseline of common understanding and approach that could be centrally governed.

Good frameworks exist for procurement, such as the Office for AI's procurement framework [3] and for ethical assessment - the ATI ethics framework [2] - and these should be deployed, perhaps with some specific tailoring. We feel there is a potential gap around a more technical framework to manage and quality assure AI technical capabilities; to promote development and testing best practice, and to define how to measure and document algorithm performance, bias, robustness, and vulnerabilities. This technical framework might therefore be more supplier-focused.

AI and Machine Learning

AI is the general field of computers replicating human intelligence. It is often taken to be synonymous with machine learning (ML), where statistical techniques are used to build models based on previous outcomes through "training" of the model with historical data as input. ML models are particularly sensitive to the quality and extent of the training data and if this is insufficient or unrepresentative they may be biased, and not robust to different scenarios. It can also be hard to explain decisions from an ML model.

Although ML is a prominent and widely represented paradigm, AI techniques are broader than ML approaches. There are multiple other types of AI – such as rule-based and plan recognition approaches – which are transparent in their reasoning, and less susceptible to bias or robustness issues. These should, of course, still fall under the same ethical and quality assurance frameworks.

A system view of AI

Although it is possible to view an AI algorithm as an "off-the-shelf" product sold by a vendor, we believe a wider view of an "AI solution" is more representative of implementation reality – that is, the AI solution includes the algorithm, the use case, the data, the governance model and the user with equal importance, and managing the solution involves both the vendor and the purchaser. In our discussion below we generally assume we are talking about the whole AI solution rather than the specific algorithm.

Another systems aspect is the AI lifecycle. There often isn't a fixed delivery of an algorithm that "just" works, but instead it may be iteratively developed, tested and tailored through trial and proof of concept stages, and will need monitoring, maintenance and review once in service, in order to identify issues and drive improvements.

Specific responses to the Committee's questions

Procurement

When a public body is purchasing a sophisticated technology, what are the key milestones of the process? What would they be in an ideal process?

In general, the delivery of an AI solution should follow government technology methodologies (e.g. GDS solution lifecycle [4]). Ideally, the key stages include requirements elicitation; concept development; ethical assessment in order to proceed; one or multiple solution options appraised with sufficient technical assurance; technology trialling (with real data); iterative improvement; user

training and user experience (UX) development; final quality assurance (QA) before "go live"; deployment; and in-service management and monitoring, including regular review of the performance of algorithms.

Based on our experience, we would emphasise the importance of some elements of this process. The understanding of the requirement (articulation of the business problem) is critical to avoid making too many assumptions about the technology solution up front, particularly when based on existing ways of working. Secondly, trialling and technology improvement stages can be difficult to estimate in terms of elapsed time as often a new business problem is being solved.

To what extent do purchasers and vendors comply with AI procurement guidelines? Where guidelines are not complied with, why do you think that is?

The Office for AI's procurement guidelines are well presented and are useful in terms of framing the key steps and risks in procurement. As BAE Systems has not been a direct AI supplier it is difficult to comment on how well they are utilised or how they are complied with. We would consider adding to the guidelines independent ethics committees and an AI "supply chain" assessment (model code components, training data, and operational data).

In general, the Office for AI's procurement guidelines are aimed at the purchaser, and they contain relatively little guidance for the vendor/supplier. This would also be an area for extending the guidelines in a policing and justice context. This could include expectations on the supplier of development and testing approaches, consistent data assessment, and how model performance characteristics are "assured" within a QA process; how the different models of procurement support different use cases; and the nature of the ongoing development and support relationship between purchaser and supplier.

What are the main hurdles to purchasing or selling new technologies for the application of the law?

One hurdle is developing a mutual expectation between purchaser and supplier regarding the AI solution. They are not normally "off-the-shelf" or commodity procurements, but instead may have varying degrees of technology maturity, and will need optimising for the use case. It is important that purchasers recognise this and can support concept stages, collaborative and joint development, and trialling and testing on real data. This also needs joint technical and domain expertise. It is helpful to have a common technical language, such as Technology Readiness Levels.

Accompanying this is the opportunity for flexible procurement models. For novel technologies such as AI there can be benefit in collaboration between purchaser and supplier (potentially at each other's cost) to develop concepts for new business problems, but without prejudicing future procurement of the operational solution which would follow normal rules. This would encourage joint innovation and would also allow effective engagement with SMEs.

Another hurdle is taking account of the broad ethical considerations of introducing a novel technology for policing and justice – that the necessity and proportionality considerations are addressed, and that the other key

considerations such as fairness and transparency or explainability are embedded from the outset. The risk of procuring a "black box" solution needs to be mitigated.

Should the procurement of advanced technologies for police use be centralised (for instance in the Home Office or within a policing body)?

We would support some form of centralised AI procurement within policing and justice. Even with the use of standards and frameworks it will be hard to embed sufficient expertise across all the police forces and justice organisations. In addition, there may be common capabilities (or multiple repeated instances of such) that could be developed and procured once. The Police Digital Service would be a candidate for the centralised role for police forces, as would the Home Office – either would ensure alignment with the Office for AI's procurement guidelines and consistent application and monitoring of procedural and other safeguards.

Transparency

The World Economic Forum recommends governments "ensure that AI decision-making is as transparent as possible". How can this be achieved?

Transparency is important to both support public trust and to ensure effective decisions can be made. It can be considered at two levels – transparency in how AI has reached a decision in any specific instance and transparency in where AI is being used for decision-making. For this question we consider the first definition.

In our experience, transparency means different things to different users. From an analysts' point of view it is about corroboration – ensuring that they can understand the AI conclusions and appreciate how they would have reached the same conclusion themselves given the same information. From a governance perspective it is about clarity of consistent behaviour across different scenarios rather than specific individual outcomes. Finally, from a citizen's perspective it is about having sufficient information to understand the application of AI and challenge an outcome.

We usually consider the issue of transparency explicitly in relation to ML-based AI approaches as this is where it is problematic; it is often hard to determine how a trained ML model is reaching a decision due to its statistical complexity. There are other AI techniques for policing and justice for which transparency is much clearer.

For ML-based approaches, achieving transparency requires trade-offs, and a decision has to be made at the design stage balancing ethical and performance considerations. There are two approaches we have previously used to maximise transparency:

1. Selecting a more transparent algorithm, for example a decision tree rather than a neural network approach
2. Using techniques that infer the rules used by the model. There are a range of such techniques highlighted by the ICO [5]. One of these – LIME (Local Interpretable Model-agnostic Explanation) – provides "local explanations" by modifying input data and observing the impact

on the model's output. We have used LIME in a non-policing application.

Deep learning, which often underpins image analytics approaches, presents real challenges as models are composed of millions of numerical weights and so are humanly uninterpretable. Some insight can be gleaned from a good understanding of design and training datasets but their use for high risk applications should be treated with caution unless expert human judgement or a LIME-type technique is viable.

What balance can be struck between transparency and commercial confidentiality?

Transparency does raise complications with respect to preserving commercial confidentiality. If transparency reveals sufficient detail on how the AI solution works then it may undermine a supplier's intellectual property (IP), which is problematic, especially when supplier internal investment has developed the core concepts. In an extreme case, full transparency may discourage suppliers from wishing to put forward technical solutions to a problem.

To date this has not become an issue. The use of a local explanations such as LIME provides insight into the model but preserves novel IP. Non-ML and simple decision tree approaches are unlikely to have substantial IP in their explainable models.

It's been suggested that this Committee recommend a public register of algorithms used in the justice system. What is your view on this? What information should be on this register?

We are very supportive of being transparent with the public regarding the use of AI in the justice system. However, we believe that that the scope needs careful consideration. In particular, we would argue that a register of all AI algorithms is in itself not useful. Such a register would be hard to maintain (what would count as an algorithm?), and may not give the public sufficient information on the application. Inclusion of low maturity algorithms in early phases of development could also stifle innovation.

Instead, we would suggest that there could be a high-level public register of the operational AI use cases in policing and justice that might impact the public. This would include summarised information on data, algorithms and usage, and could be supported by some additional ethical and technical QA information to increase public confidence. It should exclude non-operational algorithms (such as those in evaluation or research and development), those just used tactically, and those for low risk use cases (e.g. email filtering).

We would not recommend providing algorithm details, actual training data or model code in a public register as this could damage supplier IP, make solutions vulnerable to future criticism (e.g. if the algorithm is superseded by a better one), as well as susceptible to adversarial attacks or data "poisoning" from those wishing to circumvent them. Data poisoning is where an adversary inserts incorrect data into a model's training data, damaging the model's effectiveness.

Aside from the public register, we would support private registers/libraries of AI algorithms (and versions) as is best practice for AI model management (e.g. MLOps approaches).

Quality Assurance

How much information do forces ask for about how the product/solution has been developed and tested, and data around accuracy and reliability? Is this requested information in fact provided by commercial companies?

The Committee has been told that some vendors have made unproven claims about products. How could public bodies be confident in the scope, quality, and legality of the technology they are procuring?

It seems sensible to address both these questions together, and in particular to concentrate on the points about lack of consistency in understanding AI accuracy, reliability, and quality from suppliers.

We recommend that these points are addressed through clear standards and guidelines/frameworks for technical quality assurance (QA). These should be clearly communicated to suppliers with the expectation that they can respond against these standards. There should be a graded approach to support the continuum of AI development – guidelines and standards should be less onerous for low maturity and research applications, as well as being different for low and high risk applications.

We would expect the technical guidelines and standards to cover "lab" and real performance, bias assessment, commentary on robustness to new data and overall reliability, vulnerabilities and explainability of outputs. In addition, there should be a link to the ethical assessment framework and the likely use cases that the AI can be used for.

QA should be regularly revisited as the solution is developed to fit the specific operational use case. Once operational, is it important that any AI technique has regular (and ideally independent) monitoring and inspection, to ensure it is still working effectively.

There is scope for centralised ownership of a QA framework, either specifically within policing and justice (e.g. the Police Digital Service) or nationally with an independent statutory body, such as the ICO. A central governance body could undertake proactive assessment of emerging and in-service AI techniques, and could also develop a certificate of conformance (or kite mark/CE label) for approved AI applications, which would further strengthen public trust.

Public trust

Witnesses who have appeared before this Committee were concerned that the use of technology trained on biased data may affect trust in the rule of law. How could this concern be addressed?

Public trust will be impacted if model outputs are biased because the models will be perceived as being unfair on some part of the population, and hence bias management is at the heart of any ethical framework. Being able to demonstrate that a system is unbiased and can engender public trust will not be trivial.

Bias is a particular issue for ML models trained with unrepresentative data (e.g. limited ethnicity or age range) as predictions outside the population seen in

training will be inaccurate.. Bias may also occur if the training data labels have historical human biases within it. Other (non-ML) AI techniques, such as rule-based or plan recognition approaches, are significantly less prone to bias, but engineers could still encode data features that could, for example, depend on protected characteristics.

Bias can be addressed in three ways:

- Acknowledging it probably does exist in many models;
- Identifying if and how the model of interest is biased; and
- Correcting the bias to the greatest extent possible.

Bias can be assessed through assessment of the AI "supply chain" – answering questions such as where the model originated, what data it was trained on, and whether any ethical frameworks were used. For ML, it can be assessed through testing and statistical monitoring of training data and model outputs, and can be rectified by retraining the model with more representative data, labels or features. One approach is to change data sampling to ensure that all the sub-populations in the data are fully represented.

We believe it will be important to demonstrate that model (and decision) bias has been removed to the greatest extent in all policing and justice applications, and in particular those that are mission-critical or have direct public impact. This should form part of the QA process, and it should be monitored and reassessed throughout the model lifecycle, from concept to disposal.

On the positive side, it is probably easier to remove bias from a model than it is to change undesirable human behaviours, and it is a technically solvable problem that may ultimately increase public trust.

Responsibility

Who should be held accountable for the misuse or failure of technology? How long should the contractual relationship between purchasers and vendors last?

Accountability for AI technology outcomes and mistakes is one of the key considerations for a technology that could have such a significant societal impact. At a high level we believe that the organisation operating the capability needs to be ultimately accountable for its performance and correct operation, however we do recognise that detailed accountability for mistakes is probably shared between purchaser and supplier. Misuse may be caused by purchaser staff behaviour, decision bias, or poorly delivered training / business change by the supplier. "Failure" could be due to algorithm coding, training data, operational data quality, biases, explainability errors, or the wrong use case. In both scenarios it may be hard to distinguish ownership of an issue.

We recommend that the sharing of responsibility is considered on a case-by-case basis at the bid initiation / RFP stage of the project lifecycle, drawing on the ethical framework and process based governance. The relative responsibilities, risk appetite and mitigations will need to be agreed between purchaser and supplier based on specific implementation parameters.

If accountability is shared it implies that the supplier will be contractually involved with the operation of the AI solution to some extent throughout its

lifetime, so extended support, monitoring and maintenance arrangements should be considered.

Where technology is licensed, we would encourage regular renewals of maintenance and licenses – rather than one-off purchases – in order to maintain mutual commitment and foster further development.

Whose responsibility is it to train officers so they feel confident to interpret and challenge the output of an algorithm?

Officer training is also a shared responsibility. Although ultimately the need for training is owned by the customer organisation, the supplier may be best placed to conduct the required business change and provide specialised training on the AI solution outputs and their interpretation, as well as ongoing user support to provide assurance regarding training quality. It may be that the customer would provide more general skills training such as data and statistical awareness, perhaps centrally.

Training should include governance requirements and the mitigation of decision bias, supported by explainable outputs where possible. The effectiveness of training could be measured through monitoring of model outputs and user actions.

Legislation

Would you prefer soft guidelines or hard regulation?

It is important to have both guidelines and regulation. Guidelines (and frameworks) can be developed and discussed with suppliers relatively effectively, and they can be quickly iterated to improved versions. Consideration should be given to whether the various frameworks and guidelines could be drawn together into a single supplier Code of Practice.

Regulation is slower to implement but provides a concrete operating framework to ensure proper consideration of use cases, ethics, techniques and operational purpose that will build public trust. It is important that regulation does not stifle or slow innovation or discourage SMEs or academia from involvement, so it should regulate the governance of the application to the use case, rather than the development of the technique or algorithm.

We suggest regulation is probably best addressed at the national level as it would best reflect national interest and risk appetite. However, there should be some coherence and alignment of concepts with international regulation as it emerges.

We also note the potential requirement for an independent statutory AI governance body.

8 November 2021

References

[1] "(Artificial) Intelligence Led Policing", BAE Systems White Paper, December 2020

[2] “Guidelines for AI procurement”, Office for Artificial Intelligence, https://www.gov.uk/government/publications/guidelines-for-ai-procurement

[3] “Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector”, Dr David Leslie, Alan Turing Institute, June 2019, https://www.turing.ac.uk/research/publications/understanding-artificial-intelligence-ethics-and-safety

[4] GDS Service Manual, https://www.gov.uk/service-manual

[5] “Explaining decisions made with AI – Part 2: Explaining AI in practice’ ICO and Alan Turing Institute, https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/explaining-decisions-made-with-artificial-intelligence/part-2-explaining-ai-in-practice/