Artificial Intelligence Policy (Surrey and Sussex) (1236/2024)

## Abstract

This policy provides information on the use of Artificial Intelligence.

## Policy

### 1. Introduction

1.1 Artificial Intelligence (AI) is increasingly being used for its potential to bring substantial benefits to the way that services are delivered. If used safely and appropriately, AI can improve how we manage and use data and help us to communicate with and support citizens, our staff, third party partners and our suppliers more efficiently.

As such, with the emergence of new AI technologies, particularly around Generative AI, the public debate around legal and ethical implications of AI systems, as well as the negative effects they could have on society and humanity, has exploded. This policy and procedure are intended to help staff understand the Surrey Police and Sussex Police (hereafter referred to as the Forces) position on the use of AI technologies within its services and empower people to leverage artificial intelligence responsibly.

1.2 This policy and procedure are targeted at users of AI technologies and architects, developers, data scientists and security experts tasked with designing and building solutions, applications, and plugins leveraging AI related technologies.

The following should also be aware of the content of this policy and procedure, in order that they can provide appropriate oversight and governance of the use of AI related technologies within the Forces:

• Senior Information Risk Owner (SIRO).

• Information Asset Owners (IAOs).

• Information and Cyber risk practitioners and managers.

• Auditors providing assurance services to the Forces.

Additionally, the Force’s reliance on third parties means that suppliers acting as service providers or developing products or services for our use, should also be made aware of and comply with the content of this policy and procedure, in relation to their work on Forces' systems and data.

Please note that this relates to AI technologies being designed, developed, or procured by the Force, and the use of AI tools including Large Language Models (LLMs) such as Chat Generative Pre-trained Transformer (ChatGPT).

The central government response to the use and regulation of AI is still evolving. In this context, the Forces intend to remain dynamic with policy provision in this area, whilst still providing clear guidelines on how AI should be used. This policy and procedure will be adapted as necessary to developments in the national landscape and to legislation or policy introduced by central government.

**2. Scope**

2.1 In scope for this policy are data management and cyber security considerations and requirements for the:

• Acquisition and implementation of solutions that incorporate AI.

• The potential negative impacts AI may have on individuals and communities.

• Development of solutions with integrated AI.

• Use of AI tools, e.g., ChatGPT, Google Bard and Large Language Model Meta AI (LLaMA).

2.2 Out of scope are:

• Legal or Regulatory considerations, other than those directly related to cyber.

• Weaponisation and/or use of AI against the Forces, e.g., AI-generated phishing, vishing, fake profiles, malicious chatbots and advanced malware.

• Other key AI usage principles, which are covered by the National Police Chiefs' Council (NPCC) endorsed - Principles for Using Artificial Intelligence (AI) in Policing.

2.3 This policy does not standalone; it is important that all the Force's protective security procedures and the National Artificial Intelligence Cyber Standard are considered during the acquisition and development of AI-based solutions.

Example existing policies and procedures that should be consulted are:

• Data Protection Policy (Surrey and Sussex) (780)

• Business Continuity Policy (Surrey and Sussex) (1057)

• Force Information Security Policy (Surrey and Sussex) (722) and the associated procedures detailed below.

• Access to Information Procedure.

• Information Systems Through Life Assurance Procedure.

• Security Incident Management Procedure.

• Cryptography Procedure.

• Third Party Assurance Procedure.

• Configuration Management Procedure.

• Boundary Protection Procedure.

**3. Policy Statement**

3.1 Use of AI must be in a manner that promotes fairness and avoids bias to prevent discrimination and be used in such a way as to maintain security of the Force's information, comply with Data Protection, the Human Rights Act 1998, and contribute positively to the Force's goals and values.

3.2 Users may use AI for work-related purposes subject to adherence to the following procedure.

## Procedure

### 1. What is Artificial Intelligence (AI)?

1.1 Artificial Intelligence (AI) refers to computer systems capable of performing tasks that would normally require human intelligence. These systems can take many forms, and what is popularly considered as AI is continually evolving as AI technologies become more embedded in everyday human life. Some common forms of AI technology include algorithms and predictive analytics, chatbots and virtual assistants, Machine Learning (ML), remote monitoring tools, smart technologies, text editors and autocorrect, automatic language translation, and facial detection or recognition.

### 2. Definitions

2.1 There are many definitions available for Artificial Intelligence (AI). For the purposes of this document, the definition provided by the NPCC endorsed – 'Principles for Using Artificial Intelligence (AI) in Policing', written by Science & Technology in Policing – is being used. Whilst the term 'AI' is referenced throughout this document, it is intended that reference to AI, covers all of the below.

• Artificial intelligence (AI) refers to a machine that learns, generalises, or infers meaning from input, thereby reproducing, or surpassing human performance. An example is using image analysis to determine whether a video contains sexual activity with a child. The term AI can also be used loosely to describe a machine's ability to perform repetitive tasks without guidance.

• Machine learning (ML) refers to algorithms that leverage new data to improve their ability to make predictions or decisions, without having been explicitly programmed to do so. ML is a widely used form of AI that has contributed to innovations such as speech recognition and fraud detection.

• Advanced Data Analytics (ADA) uses subject matter expertise and techniques that are typically beyond those of traditional business intelligence to extract insights and make recommendations from complex data. The techniques vary widely, from data visualisation to complex linear models to language analytics. An example is the use of Risk Terrain Modelling to quantify environmental factors that shape risk mapping and resource deployments.

There is other related AI terminology, which start to overlap with the above, but are included here for completeness. The above and below are considered the most common, but there are others. The additional definitions are:

• Generative Artificial Intelligence (GAI) is artificial intelligence capable of generating text, images, or other media, using generative models. Generative AI models learn the patterns and structure of their input training data and then generate new data that has similar characteristics.

• Large Language Models (LLMs) are a subset of GAI, where an algorithm has been trained on a large amount of text-based data, typically scraped from the open internet, and so covers web pages and, depending on the LLM, other sources such as scientific research, books, or social media posts. Examples include ChatGPT, Google Bard and Meta's LLaMA.

• Natural Language Processing is a computer's attempt to "understand" spoken or written language. It must dissect and analyse vocabulary, grammar, and intent, and allow for variation in language use. The process often involves machine learning.

**3. How Do I Know If My Technology / Project Is Using AI?**

3.1 For some technologies, it is usually obvious that they operate using AI, however this is not always the case. If you are unsure if a technology you are using, or plan to use, would be considered as AI, it may be helpful to consider the following:

• Does it support decision-making or make decisions?

• Does it support the delivery of information?

• Does it autonomously identify patterns in large volumes of data?

• Does it utilise Machine Learning, for example, learning to answer questions or solve problems?

• Does it predict or manage risks?

• Does it contribute to the allocation of resources or prioritisation of actions or investigations?

• Does it translate language?

• Does it analyse and/or act on data from its environment?

• Does it perceive and react to the world, for example, recognising visual information (e.g., objects, individuals) or speech?

• Does it store past data and predictions to inform future predictions?

• Does it remember, adapt, or encourage changes to behaviour patterns?

3.2 If the answer to one or a few of the above is yes, then it is likely that the technology is using AI to operate, and you will therefore need to follow the requirements and considerations set out within this policy.

**4. What Are The Risks Associated With The Use Of AI?**

4.1 Whilst understanding of the risks associated with the use of AI is still developing, some of the key risk areas that have been identified in research and practice thus far include:

• Data Protection.

• Security Vulnerabilities.

• Transparency.

• Bias.

• Automated Decision Making.

4.2 In addition, there are some specific high-risk AI technologies that individuals should be aware of, for example, chatbots, and ChatGPT or other Large Language Models (LLMs).

**5. Data Protection Considerations**

5.1 There is currently no legislation in place that directly refers to the use of AI. However, where an AI system is using or collecting personal data, it will fall within the scope of the UK General Data Protection Regulation (UK GDPR) and the Data Protection Act 2018 (DPA 2018). This could include where personal data is being used to train or test AI, and/or in the deployment of the technology. The regulation grants individuals' certain rights where their personal data is being used or created, particularly for automated decision making. These rights must be considered in the development and use of all relevant AI technologies, so you will need to review and consider the implications of this for your specific project or activity; as such, if personal data is being processed you must complete a Data Protection Impact Assessment (DPIA) DPIA Questionnaire. Note that if the data is categorised as 'Special Category Data' under the UK GDPR or as 'Sensitive' under Part 3 of the 2018 Act (Law Enforcement), a DPIA must be completed together with an appropriate policy Document. 'Sensitive' data, which will include all biometric data, should only be processed if strictly necessary. The advice of the respective Force Data Protection Officer (redacted text) should be sought in each case.

5.2 If you are considering the development of any technology which involves the processing of personal data, consideration should also be given to ensuring that the data rights of subjects can be met. For instance, is there likely to be an increase in Data Subject Access Requests / Right of Access Requests or Freedom of Information (FOI) requests and do the Forces have capacity? Will the data be easily retrievable and able to be provided to the data subject when requested? Retention and disposal schedules should also be considered. Development should always consider the general principle of Data Protection by Design and by Default (Art 25 UK GDPR).

**6. Security Vulnerabilities**

6.1 When embarking on an AI related project always follow a 'Secure by Design' (SbD) methodology and ensure:

• There is appropriate logging and monitoring requirements for each AI instance

• There are access controls in place to limit who can input or amend data within the AI solution, and alert when unauthorised or unexpected changes are made

• There are vulnerability management tools and techniques deployed to identify at risk AI systems and harden them against compromise

• Threat intelligence feeds to watch for attacks or compromises of AI tools are updated and maintained

• Suppliers or open sources meet existing security standards on ID management, access control and authentication

• Supply chain assurance extends to software and service providers who help to develop and maintain organisational AI systems

• The AI solution has the appropriate levels of uptime and resiliency, and that the technology platforms underlying the AI system also meet the levels of uptime and resiliency required

• Any integrated or accessible AI should have a clearly defined tested and assured process to immediately cease any active or ongoing processing or tasking that the AI is undertaking to mitigate any operational or platform impacts on live services should said processing go beyond expectations

• There is an appropriate level of network segregation and access control in place to limit access to the AI system via lateral movement and direct access

• Red team testing (a cyber-attack technique) is deployed to explore possible attack patterns

• The information the AI model returns to queries is managed to limit the ability of threat actors to gather useful attack data

• The solution is recorded in the relevant asset inventory, with a clear description of its approved use, who owns it and other data relevant to the AI system and its use.

6.2 The Force's Security Operations Team and Digital, Data and Technology (DDaT) technicians must not become over-reliant on the decisions of an AI based security system and therefore should consider:

• Establishing guardrails or dashboards for what normal security operations looks like and regularly audit to gauge whether the system is operating as expected.

• Establishing a rapid reaction process to examine and investigate exceptions and anomalies to determine causes.

• Keeping the Security Analysts and Technicians 'in the loop' – build processes and procedures around the AI based security solutions, that ensure there are suitable human checkpoints to verify and act on any recommendations.

**7. Transparency**

7.1 It is also good practice to maintain the principle of transparency in the use of AI throughout the process as well as the capability to understand why an AI system reached

a particular decision, recommendation, or prediction. This means, establishing a clear understanding of the purpose of the technology from the outset; establishing individual responsibility and accountabilities; ensuring that operational staff and senior managers have a good understanding of how the AI operates, and ensuring, if relevant, citizens are aware of the use of AI and what it means for them.

**8. Bias**

8.1 For AI technologies data is crucial, particularly with regard to the potential for bias and discrimination. Users must understand the equality and human rights implications of AI so it is important that the data sources that will be used to train AI, as well as the data sources that the technology will be using to make its analysis or predictions, are assessed for potential unconscious bias or discriminatory outcomes at the start of an AI project – all users, Surrey Police and Sussex Police, should consider the principles set out in the Sussex Police Data Analytics Strategy when making their assessment. It is a good idea to engage a diverse team with a variety of perspectives to undertake this exercise to ensure all potential discrimination or bias is identified; this could include staff groups, stakeholders, or service users. Additionally, the datasets must be subject to extensive testing and sampling to expose systemic, computational, or human cognitive bias. Depending on the scale of the project or activity, it may take some time for trends indicating bias to become evident; so continued output monitoring is important as well as consideration of existing research into the use of specific AI technologies which might offer advance warning of the bias problem. If potential bias is identified, an alternative data source may be selected, or the design of the algorithm will need to be tweaked in terms of how it functions and makes predictions.

8.2 When selecting the data source(s) that are to be used (to train the AI, or to be processed by the AI), you must consider whether there could be any bias in how the data set was collected; for example, stop and search data or arrest data is not collected in a way that is free of bias because a human has been able to apply discretion in their decision-making. Similarly, you will need to consider data quality and type. This will involve considering if the data is complete and accurate:

• Are there are any gaps in protected characteristic information?

• Is the data sufficient for trend identification?

• Is protected characteristic information self-reported?

• Who owns the data?

• Does the data reflect the group of people who are the intended audience or users?

If the data quality is poor, you will need to invest in improving data collection before proceeding to develop an AI solution.

• Do the data sources include personal data or commercially sensitive information? You must conduct a DPIA and seek advice from the respective Force Data Protection Officer (redacted text) to ensure any sensitive data is shared appropriately and legally.

8.3 The outputs of AI can also be impacted by the human decisions made in its design – the selection of data used to train it, the assumptions that inform the algorithm, and the

way in which its outputs are interpreted and applied. Therefore, there is the potential for the AI to perpetuate existing bias or inequalities. Whilst this can be mitigated by making a considered choice when selecting the most appropriate data to use, steps also need to be taken to mitigate any assumptions embedded within the algorithm itself. This will involve:

• Using an Equality Impact Assessment (EIA) to conduct a robust assessment of the existing processes or current practice that is proposed to be supported or replaced by AI, before AI development or procurement commences. You need to consider if there is potential that there is already embedded unconscious bias or discrimination occurring that will specifically need to be addressed in the design of the algorithm.

• Where relevant, utilising the principles of inclusive design to involve people who will be affected by the technology, to ensure that the AI's assumptions or outputs take into account their experience.

• Devising a methodology to monitor the actual impacts and validate the AI's outputs. It is a good idea to consider how any assumptions might be impacting on the AI's outputs, as algorithms will attempt to match previous predicted behaviours to outcomes and thereby reflect the expectations of the humans designing it.

• Reviewing the outputs and considering why certain protected characteristics are being identified more or less than others.

**9. Automated Decision Making**

9.1 AI will not necessarily produce perfect and accurate predictions or outputs. Algorithms can yield false negatives or positives, which can reproduce bias or inequality. As such, the risks can be even greater where an algorithm is supporting human decision-making or resulting in automated decision-making, because this could potentially allow incorrect or biased outputs to be implemented unchecked. AI solutions expose little information about how decisions are reached, and the Forces could be exposed to regulatory sanctions, if explanations or justifications are absent for decisions made. Therefore, the individuals responsible for implementing or using AI solutions must:

• Follow a 'Secure by Design' (SbD) methodology (see section 6.1).

• Test extensively to gain as much understanding as possible about what outputs are likely, including against different user or citizen groups to look for bias.

• Run parallel models to explore how decisions and outputs can change.

• Talk to regulators about their expectations to set levels for transparency and explainability.

• Prove and demonstrate the efficacy of the entire system.

If the AI solution has been acquired, then the Forces must seek clarification from the provider that the above has been carried out as part of the procurement exercise.

9.2 Steps must be taken to ensure that human challenge and oversight is retained in all use of AI. This is important because it allows for any errors to be identified, can prevent discriminatory outcomes, and provide opportunity for bias to be identified and addressed.

To support the function of human challenge, staff using the technology must have an understanding of how the algorithm operates and be provided with additional training as required, such that they are fully equipped to identify any errors.

**10. High-Risk AI Technologies Chatbots**

10.1 Whilst chatbots can vary widely in their specific capabilities and complexity, chatbots can be broadly defined as computer programs that simulate and process human conversation in their response to questions received from a real person. Some more sophisticated chatbots such as Apple's Siri, Google Assistant and Amazon Alexa, are now more commonly referred to as 'virtual assistants' or 'virtual agents'.

10.2 There is a specific range of risks associated with the use of chatbots, arising from the fact that chatbots interact with users or citizens, rather than operating 'behind the scenes'. As such, it is important that sufficient consideration, and where relevant, mitigations, are in place to protect the intended users. Some key considerations include

• It must be made very clear to users that they are speaking with a chatbot and not a human so that they can make an informed choice as to whether to continue the interaction or not.

• A human-based alternative must be made available and easily accessible should the user choose to opt out of engaging with the AI, or should they struggle to have their needs met by the chatbot.

• Where a chatbot is to be used by children or might be accessible to children (or other vulnerable user groups), the potential safeguarding risks need to be adequately considered. This encompasses both the need to ensure that the chatbot is not giving harmful advice, and the need for the chatbot to recognise certain information that might be provided by a user, indicating that they are at risk or in danger.

• As with other forms of AI, the functionality of the chatbot will be dependent on the quality of the data used to train it. Chatbots can cause bias toward certain users if not designed or programmed properly. To mitigate this, it is important that the chatbot is trained on data that is accurately representative of the groups that will be using it.

**11. High-Risk AI Technologies ChatGPT and Large Language Models**

11.1 ChatGPT is a Large Language Model (LLM) chatbot developed by OpenAI. It uses Deep Learning technology to provide human-like answers to questions asked by users. Whilst this type of AI has considerable potential capabilities, it also carries significant risks, which means that all use of these technologies must be conducted in a safe, appropriate, and accountable manner. When and where available, individuals are expected to make use of LLM technologies (and other AI tools) provided via the Microsoft suite of applications, as the primary option in place of other alternatives.

11.2 The following summarises some of the associated risks that staff will need to consider and assess if thinking about using ChatGPT or other LLMs in their work:

• Individuals must not input personal, operational, or sensitive police data, without prior consultation with the respective Force Information Management Team, a full understanding of the risks by completing a DPIA, and sign-off from the appropriate risk

owner. If individuals believe that they may have input such data without undertaking a DPIA, they must follow the Forces Data Breach Procedure. Security and Breach Reporting.

• ChatGPT and other LLM technologies can provide answers that are superficially plausible, but potentially incorrect. For this reason, they must not be used to support operational decision making or in the creation of documents that may enter the Criminal Justice system, without prior consultation with the appropriate operational process owners, Legal and Information Management Teams, a full understanding of the risks, and sign-off from the appropriate risk owner.

• Information inserted into ChatGPT is not confidential and could inadvertently end up in the public domain. Users must disable the chat history to ensure the query information provided does not become part of its future training dataset.

• As with other AI technologies, there is the risk of bias and the production of discriminatory answers. This is exacerbated by LLM technologies that have an extensive data source which makes it impossible to completely filter out offensive or discriminatory content.

• The breadth and extent of the internet data that LLMs are trained on is also likely to include copyrighted material, with answers generated without any source references, which could pose a potential Intellectual Property or copyright issue for its outputs.

11.3 In addition, staff should be aware that other malicious forces are likely to make use of LLMs to exploit any Force vulnerabilities, whether this be for the development of more sophisticated hacking techniques, or the production of more convincing phishing emails.

11.4 The use of AI for completing assignments, essays and exam responses is strictly prohibited, such actions compromise the learning process and violates the principles of originality, honesty and integrity.

**12. Procurement**

12.1 It is most likely that the AI technology used or acquired by the Forces will be designed and developed externally by a third party, and therefore most AI or AI-related projects will likely include a commissioning or procurement exercise. Individuals should continue to follow the Force's existing policy and guidelines with regard to commissioning and procurement but will need to be aware of the unique challenges associated with AI technologies. This means working with the supplier to fully understand the risks and considerations that have been made in the AI's development, as ultimately responsibility for the outputs will sit with the Forces. Individuals should also ensure they undertake due diligence when selecting a supplier, even if the supplier pool is small, and should therefore carefully consider the risks and limitations of this technology.

12.2 There are a number of interdependencies that will need to be managed between the Forces and the supplier, including:

• The ownership of the data that the AI is trained on.

• Transparency regarding the design and assumptions of the algorithm, the extent to which this can be shared between customer and supplier.

• Understanding of the legal and ethical accountabilities.

• Responsibility and capability for oversight of the technology, monitoring and potential rights around requesting changes.

• Integration into existing Force processes.

**13. Governance Framework**

13.1 The Chief Digital and Information Officer will take the lead with regard to setting the policy and standards for the use of AI, and the Force Security and Information Management Board will take responsibility for ensuring the standards set out in this policy are adhered to. As part of its governance responsibilities for AI, this Board will ensure the Forces:

• Has the requisite skilled people, process and technology elements required to set up, run and maintain operational AI systems and cope with their outputs.

• Maintains awareness of the evolving legal and regulatory changes impacting the use of and security of AI, on an ongoing basis.

• Monitors guidance from regulators on their standards for datasets.

• Adopts a risk-based, rather than compliance-centric, approach to help the Security and Information Assurance Team and other appropriate stakeholders meet regulatory and legal requirements, but not stifle innovation and the potential benefits of the use of AI in policing.

• Undertakes a regular policy gap analysis to address shortcomings as AI technology evolves.

• Reviews Incident Response Plans and ensures potential AI related incidents are adequately covered and there are Incident Playbooks which are regularly tested, e.g., personally identifiable data being leaked.

• Maintains a register of how we use AI (what systems, what decisions as a result, how we assessed the bias, what algorithms are used etc.) and the associated documents e.g., DPIA, EIA, for the purposes of being able to respond to FOI requests.

13.2 Note also that as part of its governance role for new technology solutions, the IT Architecture Review Board will be scrutinising technology solutions against this policy and procedure and national standards to ensure compliance.

**Team:** Digital Data and Technology