September 2023

KCC Artificial Intelligence (AI) Policy

Contents

1. Introduction & Context

1.1 Artificial Intelligence (AI) is increasingly being used across industries, including the public sector, for its potential to bring substantial benefits to the way that services are delivered. KCC has also begun to use this technology in various shapes and forms. If used safely and appropriately, AI could improve how we manage and use data and help us to communicate with and support residents, service users and suppliers more efficiently. Understandably, the development and implementation of AI technologies has recently received significant press attention, particularly where problems have materialised. As such, with the emergence of new AI technologies and the associated equalities and data protection risks, this policy is intended to help staff understand the Council's position on the use of AI technologies within its services. **All staff currently using AI, or intending to use AI, must familiarise themselves with this policy and have a responsibility to maintain transparency in its use.** Please note that this relates to AI technologies being designed, developed or procured by KCC officers, and the use of AI tools including Large Language Models (LLMs) such as ChatGPT. This does not include the use of existing AI technologies available via KCC's Microsoft licence, such as predictive text capabilities in MS Teams, MS Outlook or MS Word.

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

1.3 **This policy is intended to supplement existing KCC policies that are in place; as such, officers utilising AI technologies are still expected to follow existing Council policies with regard to Information Governance, Data Protection, development of technology projects, ICT Compliance and Risk, and Equality Impact Assessments (EqIAs), as detailed below.**

2. What is Artificial Intelligence (AI)?

2.1 **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.

How can I tell if my technology/project is/using AI?

2.2 For some technologies, it is fairly 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/investigations/inspections?
- Does it remotely monitor the well-being of individuals?
- Does it predict health problems at an early stage?
- 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?

2.3 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 procedures and considerations set out within this policy. If you are still unsure, you can get in touch with one of the contacts listed at section 4.8 of this policy.

3. What are the risks of using AI?

3.1 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:

- Information Governance (IG) & Data Protection
- Transparency
- Equalities
  - Data Bias & Data Quality
  - Design Bias in Algorithms
- Automated Decision Making

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

Information Governance (IG) & Data Protection

3.3 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 **General Data Protection Regulation (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. This 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. Depending on the specific project or task, you may be required to complete a **Data Protection Impact Assessment (DPIA)**. In the first instance, a DPIA screening will be a useful tool with which to risk assess your proposed use of data; done at an early stage, this can help you to mitigate many of the potential risks associated with data protection and GDPR, especially regarding the proposed use of innovative AI technology. For more information on DPIAs and where it is mandatory to complete them, refer to the Data Protection & GDPR page on KNet.

3.4 In addition to the use of personal data, officers intending to use AI technologies also need to consider the risks associated with the use of commercially sensitive data. If commercially sensitive data, or data that we wouldn't otherwise release under Freedom of Information (FOI) requests is inputted into a Large Language Model (LLM), such as ChatGPT, or processed in another AI technology, it may have then been inadvertently put into the public domain. Releasing such information provided to the council in confidence could lead to legal proceedings. As such, it is vital that officers appropriately consider the information they are using within AI projects or activities.

3.5 To mitigate the risks outlined in 3.3-3.4, it is KCC policy that all officers proposing to use personal data, and/or commercially sensitive data in an AI project or activity must complete a DPIA and EqIA and follow relevant processes with regard to ICT Compliance and Risk. For use of an individual AI tool, such as ChatGPT, this will mean contacting ICTComplianceandRisk@kent.gov.uk , describing your proposed use and the business reason. For the development of an AI-related project, this will

mean following the ICT Commissioning Process and contacting the Technology Business Partners team in the first instance.

3.6 In addition, all projects must follow KCC's existing Information Governance policies and procedures, which are currently being refreshed in the context of AI. Other related policies and procedures, such as Data Security or Information Sharing may also be relevant to your project. For more information on these, visit the Information Governance page on KNet.

3.7 For further guidance on how to implement transparency and data protection measures within your AI project, the Information Commissioner's Office (ICO) website provides comprehensive guidance on the application of UK GDPR to the use of information in AI systems. Central government's Data Ethics Framework may also help with the planning and design of data use within your project.

Transparency

3.8 Notwithstanding the requirements of the GDPR and DPA regulations, it is also generally good practice to maintain the principle of transparency, and explainability in the use of AI, throughout the process. This means, establishing a clear understanding of the purpose of the technology from the outset; establishing officer responsibility and accountabilities; ensuring that operational staff and senior managers have a good understanding of how the AI operates, and ensuring service users are aware of the use of AI. You must take care when devising how you will communicate with service users, as they will need to be aware of the AI and what it means for them but may not have an understanding of what AI means.

3.9 For AI-related projects that will involve a greater level of interaction with the public, or have a potential for a significant impact on people, it may be necessary to complete the Algorithmic Transparency Recording Standard, which has been designed by central government to assist public sector organisations provide clear information about the algorithmic tools they use, and why they're using them.

Equalities

3.10 As a public authority, KCC must comply with the **Public Sector Equality Duty (PSED)** under the Equality Act 2010. This means that, as with any other KCC project or activity, when developing, using or procuring AI technologies, the council needs to consider the potential impact on people with protected characteristics. This consideration must be made and evidenced through an **Equality Impact Assessment (EqIA)**, conducted via the EqIA App. For more information on EqIAs, consult the Equality Impact Assessment Policy.

3.11 Some of the risks associated with AI have specific implications for equalities considerations. Typically, these are associated with the amplification of existing biases via the speed and scale of AI technologies. Such examples often receive significant press attention, can have pronounced negative impacts on protected groups, and also present a real risk of legal challenge. Therefore, it is important that these are adequately considered in the development of AI technologies at KCC. The following sections discuss this in more detail.

Data Bias & Data Quality

3.12 As the foundation of AI technologies, data is incredibly important, particularly with regard the potential for bias and discrimination. Generally speaking, as a reflection of the real world, all data has the potential to reflect current and historical structural inequalities or bias. With the addition of AI, these inequalities can then become replicated and amplified in its outputs. As such, 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. 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. Depending on the scale of your project or activity, it may take some time for trends indicating bias to become evident; this is why continued output monitoring is important. If you do identify potential bias, you may choose to select an alternative data source, or use this to inform the design of the algorithm in terms of how it functions and makes predictions. Here, transparency becomes incredibly important – the better we understand how an algorithm works, the easier it becomes to identify what is causing bias and train it out. In addition, officers should be aware of and look out for proxy variables in their data. These are variables that may appear to have a correlation, but one that is not itself directly relevant. In some cases, these can cause negative equality impacts, but once identified can be addressed in an algorithm. If the algorithms are the intellectual property of an external provider, please refer to the procurement section in this policy on page 10.

3.13 When selecting the data source(s) that you will use (to train the AI, or to be processed by the AI), you will need to consider data quality and type. This will involve considering if the data is complete – 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 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. Finally, do your data sources include personal data or commercially sensitive information? You will need to undertake the relevant impact assessments mentioned in 3.5 and seek advice to ensure any sensitive data is shared appropriately and legally.

3.14 Central Government have established the Data Ethics Framework which can be used to design and plan the appropriate use of data in the public sector, and encompasses the principles of transparency, accountability and fairness (related to data bias). It may be useful to complete the editable template available on their website for your project.

Design Bias in Algorithms

3.15 In addition to the impact of data bias, the outputs of AI are also heavily 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.[1] Therefore there is ample opportunity for 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, that adequately reflects service user demographics, steps also need to be taken to mitigate any assumptions embedded within the algorithm itself. This will involve:

- Using an EqIA to conduct a robust assessment of your existing processes or current practice that is proposed to be supported or replaced by AI, before you commence the AI development. 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, utilise 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.
- Devise a methodology to monitor the actual impacts and validate the AI's outputs. It is a good idea to consider how your assumptions might be impacting on the AI's outputs, as algorithms will attempt to 'match previous predicted behaviours to outcomes'[2], and thereby reflect the expectations of the humans designing it. Review the outputs and consider why certain protected characteristics are being identified more or less than others.

Automated Decision Making

3.16 Evidently, AI does not necessarily produce perfect and accurate predictions or outputs. Algorithms can in fact yield false negatives or positives, which can reproduce bias or inequality. As such, the risks are even greater where an algorithm is supporting human decision-making or resulting in automated decision-making, because this could allow potentially incorrect or biased outputs to be implemented unchecked. In addition, as mentioned in 3.2, where AI results in automated decision-making which has a legal or significant effect, there are additional requirements under Article 22 of the GDPR.[3]

3.17 Therefore, 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 and vetted; can prevent discriminatory outcomes and provide opportunity for bias to be identified and addressed, supporting the evolution of the algorithm. 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 errors.

3.18 It is essential that officers fully consider the appropriateness of assisted decision-making to the purpose of their project or service and consider alternatives before proceeding. If a decision is made to proceed with assisted decision-making, you will need to ensure that:

---

[1] Understanding algorithmic bias and how to build trust in AI: PwC

[2] AI & Equality Initiative: Algorithmic Bias & the Ethical Implications | Carnegie Council for Ethics in International Affairs

[3] Ethics, Transparency and Accountability Framework for Automated Decision-Making - GOV.UK (www.gov.uk)

- All of the requirements arising from Article 22 of the GDPR are met.
- A risk assessment of the use of automated decision-making is conducted.
- A responsible officer has been identified for the decisions that will be made, with their details made clear within the organisation, and for individuals who may be impacted by the decisions made.
- Consideration has been given to how the AI will fit into existing processes, and mechanisms for flagging any potentially incorrect or biased outputs has been established.
- A process for scrutiny or audit of the outputs is in place.
- Any other required mitigations are introduced.
- Central government's Ethics, Transparency and Accountability Framework for Automated Decision-Making is utilised in the design, implementation and management/monitoring.

Types of AI Technology to be aware of

Chatbots

3.19 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.[4] 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'.

3.20 There is a specific range of risks associated with the use of chatbots, arising from the fact that chatbots interact with members of the public, 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:

- Transparency is still critical; 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 from the page hosting the chatbot, 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/in danger. The Unicef Safer Chatbots Implementation Guide may be useful in considering or mitigating risk in this area. You can also use the Safeguarding at KCC page on KNet to find further information.
- 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/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.

[4] What is a chatbot? | IBM

ChatGPT and Large Language Models

3.21 **ChatGPT** (which stands for Chat Generative Pre-trained Transformer) 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. LLMs are a specific type of AI algorithm that are trained on a large amount of text-based data from the open internet. 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, officers are expected to make use of LLM technologies (and other AI tools) provided via the Microsoft suite of applications available to KCC, as the primary option in place of other alternatives.

3.22 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:

- ChatGPT has been developed by a US tech start-up, and is therefore outside of EU data protection legislation that the Council must follow. **Council officers must not input personal data or commercially sensitive information into ChatGPT before completing a DPIA, EqIA and following the ICT Compliance and Risk assessment process (see 3.5 for more information on this) to understand and mitigate the potential risks. If staff believe that they may have input such data without undertaking impact assessments, they must follow KCC's Data Breach Policy.**
- ChatGPT and other LLM technologies can provide answers that are superficially plausible, but incorrect.
- Information inserted into ChatGPT is not confidential. If chat history is not disabled, query information provided to may 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 of offensive or discriminatory content.
- The breadth and extent of the internet data that ChatGPT is trained on is also likely to include copyrighted material, with answers generated without any source references, which poses a potential Intellectual Property or copyright issue for its outputs.
- Whilst LLMs do not currently use the information submitted in queries to develop future answers/responses, this information is available to the AI providers, and can therefore be expected to be used for future model training. In addition to this, there is the risk that the AI provider itself could be hacked in the future, and thereby made publicly available. Finally, there is the possibility that an LLM system is acquired by different organisations in the future, with different terms of use and privacy policies which might put information submitted into the AI system at greater risk. **For these reasons, it is imperative that no sensitive Council information is supplied without proper consideration of the potential risks**.

3.23 In addition, staff should be aware that other malignant forces are likely to make use of LLMs to exploit KCC's vulnerabilities, whether this be for the development of more sophisticated hacking techniques, or the production of more convincing phishing emails.

3.24 Due to these significant risks associated with the use of ChatGPT, or any other LLM, **Council officers are expected to make additional considerations before proceeding with their use.** This includes:

- Considering if the proposed use is acceptable and appropriate. What are the risks? What are the actual potential benefits? How will outputs be validated?
- What are you trying to solve or achieve via the use of the technology? It should not be assumed that AI, and LLMs specifically offer the solution to every challenge. Consider if an alternative solution would be more appropriate. You can speak to the Technology Business Partners team to explore possible alternative solutions.
- Where you will be entering personal data, commercially sensitive data, or data that would not be released under FOI, it is essential that you complete a DPIA, EqIA and CaRT assessment before proceeding.

Digital Exclusion

3.25 Whilst less applicable to AI technologies used internally by staff only, the impact of digital exclusion may be relevant to technologies intended to be used by customers or service users. Mitigation of the impact of digital exclusion is part of KCC's corporate equality objectives; it also has the potential to have significant impact on individuals that experience it. For relevant technologies, this should be considered as part of the EqIA process to ensure that alternatives are in place for those who experience digital exclusion and are therefore unable to access the benefits of the AI technology.

4. What does this mean for staff?

4.1 To summarise, here are the key points to remember for staff using, or planning to use AI technologies:

- You must be transparent about your use of AI, both to service users and critically, in the completion of EqIAs, DPIAs and other project or activity documentation.
- You must be transparent about the use of ChatGPT, or other LLM programs in your work. Where you intend to input personal data or commercially sensitive information, you must first complete a DPIA, EqIA and complete the ICT compliance and risk assessment process. If you have already proceeded with use before completing these, you must urgently contact the Information Resilience and Transparency Team.
- You must follow the requirements of this policy alongside existing policies that are in place, including Information Governance, Data Protection, and the EqIA Policy.
- Speak to the contacts listed below to ensure you are getting appropriate assurance on the various elements of your project or activity (equalities, risk management, information governance, technology compliance and risk, procurement, and HR).
- Do your research. You need to make sure that you (and your staff who will be using the AI) understand the technology being used.
- Engage the right people in the development of your project or activity to mitigate any risk associated with the proposed use of AI technology.

- Start considering the potential risk factors at the earliest stage. Not only is the use of AI a significant financial investment, it can also have marked negative equalities or data protection impacts that could cause reputational damage and be costly to reverse.
- Retain the principle of proportionality. When making your considerations, consider the purpose, the users, and the specific relevance to each of the potential areas of risk.
- Make specific consideration where commissioning or procuring AI. Further detail is provided on this in the following section.
- Wherever possible, officers are expected to make use of the AI tools available within Office 365 as part of KCC's Microsoft licence as the first route before considering alternative options.
- Where the use of AI could have an impact on staff or require staff to work differently, you must contact HR for advice.

Procurement

4.1 The growth and development of AI technology has a breadth of potential implications for KCC's procurement activities, both where officers are looking to specifically procure an AI technology, and where they are not.

4.2 Where officers are pursuing an AI specific project, it is likely that the AI technology 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. Staff should continue to follow the council'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. As a burgeoning field, the AI marketplace is not yet fully developed, meaning that in some areas, available technologies may still be in pilot stage and the risks or limitations not fully understood. You will need to work with your 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 Council. Officers should also ensure that they undertake due diligence when selecting a supplier, even if the supplier pool is small. When developing such a project, officers should keep in mind that AI cannot be assumed to be the default solution to the emerging needs and challenges that are faced, and should therefore carefully consider the risks and limitations of this technology.

4.3 There are a number of interdependencies that will need to be managed between the council 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 for/rights around requesting changes.[5]
- Integration into existing council processes.

[5] Review into bias in algorithmic decision-making (publishing.service.gov.uk)

4.4 The level of consideration of these factors should be tailored to the specific project, the extent to which it is a council-designed technology (as opposed to an already existing product or one to be designed externally), and whether the entire contract concerns AI or if it is a small part of a wider contract. Central government has produced extensive guidelines that will be helpful for any officers procuring an external AI technology.

4.5 As mentioned, the risks arising from AI are not limited to the procurement of AI technologies. All staff involved in procurement processes need to be mindful that some potential suppliers are likely to be taking advantage of AI technology to develop their bids. This could lead to seemingly credible and believable content being produced as a part of the procurement process, and underlines the importance of robust due diligence, evaluation, and selection of suppliers.

4.6 In addition, AI technologies are increasingly embedded within wider services that are not exclusively AI; this could be the case for many of the services that officers are procuring or contracted for. As such, it is essential that officers utilise market and supplier engagement to understand how AI may be used as a part of the service or contract and make any necessary considerations as appropriate.

4.7 A robust approach should be taken when considering all of the above factors; comprehensive guidance on the commissioning and procurement process can be found in the How To Buy Anything pages on KNet.

Seeking Assurance

4.8 Individual services are responsible for making the considerations required of their specific AI project or AI-related project or activity and complying with this policy, and other existing KCC policies. However, if you will be developing or using AI technology, you will need to seek assurance from the following council contacts at the relevant stages of the process:

- EqIA Policy - Laura McPherson laura.mcpherson@kent.gov.uk
- Risk and Assurance – Mark Scrivener mark.scrivener@kent.gov.uk
- Information Governance (IG) – InformationGovernance@kent.gov.uk
- ICT Compliance and Risk Team (CaRT) – ICTComplianceandRisk@kent.gov.uk
- Technology Business Partners team
- Procurement - commercialstandards@kent.gov.uk
- HR - HRTeam@kent.gov.uk

4.9 In the event of disagreement regarding the level of risk identified via DPIA, EqIA and ICT Compliance and Risk assessment processes, KCC's Corporate Information Governance Group (CIGG) will act as arbitrator and make the final decision.