

LONDON BOROUGH OF HAMMERSMITH & FULHAM

Report to: Policy and Oversight Board

Date: 25/11/2024

Subject: Update on AI Adoption and Governance

Report author: Anthony King, Enterprise Architect
Ben Savage, Head of Information Management & DPO

Responsible Director: Sukvinder Kalsi, Executive Director of Finance and
Corporate Services

SUMMARY

This paper provides an update on artificial intelligence (AI) adoption at H&F and on AI use cases in other Local Authorities, and shares H&F's draft governance framework for AI.

RECOMMENDATIONS

1. To note and comment on the update.
-

Background Papers Used in Preparing This Report

Link to the Briefing paper on Generative AI in March 2024:

[POB Paper March 2024 - Generative AI: Opportunities, Challenges and Risks](#)

DETAILED ANALYSIS

1. As part of the AI governance process, the council is looking at its approach to an ethics framework for reviewing major programmes, projects or initiatives that have AI implications and impact residents and continues to monitor what other organisations are doing around adoption in the sector and beyond. Public trust is essential for the successful integration of AI technologies into the council, and this is in line with best practice employed across the public sector.
2. At the national level, we have seen the commissioning of the AI Opportunities Plan and establishment of an AI Opportunities Unit by the Department for Science, Innovation and Technology (DSIT) to develop a roadmap for government to capture the opportunities of AI to enhance growth and productivity and create tangible benefits for UK citizens. This initiative is part of

the Government's broader strategy to foster economic growth and improve public services through AI.

3. The ongoing integration of AI capabilities into commercial off the shelf applications and services has led to a shift from a technical conversation to a more business and outcome driven one. In the H&F context, this has led to services looking into how they can leverage AI capabilities to support business outcomes.
4. The Generative AI Working Group has evolved following feedback from POB, which recommended a more formal structure and means of engaging services across the organisation. H&F's Digital Strategy lead and the Chief Digital Officer currently lead this work.

Definitions

5. Generative AI is a branch of artificial intelligence that can create new content or data based on existing examples or rules. Traditional AI uses symbols and rules to show and change information and uses search methods to find answers in a fixed set of options. Traditional AI is a chess program that uses predefined rules and strategies to choose the best move. Unlike Generative AI, Traditional AI program does not learn from its own experience or adapt to new situations but follows the fixed rules and algorithms programmed into it.

AI Governance framework

6. Appendix 1 outlines a draft AI governance framework for addressing the use of AI in programmes and projects. Key Components of the governance framework process are as follows:
 - AI governance efforts are not standalone but integrated into the wider governance structure. This ensures consistency and coherence across all programmes, projects, and initiatives, enabling better oversight and visibility of AI capabilities within H&F.
 - As part of the governance framework, we have established protocols for data privacy, ethics, and information security through the information management governance process and the technical review processes such as the Technical Design Authority (TDA) and enterprise architecture framework.
 - A myriad of products with AI capabilities are growing at pace and this has meant that we have had to ensure that our AI governance processes extend to third party vendors and solutions.
 - We will include a verification process in the AI governance framework to ensure the safe adoption of AI-enabled commercial off- the shelf products when these tools and capabilities are implemented without direct input from H&F's Digital services.

7. As part of AI governance process, the council is looking at its approach to an ethics framework for reviewing and major programmes, projects or initiatives that have AI capabilities implications and impact residents. Public trust is essential for the successful integration of AI technologies into the council, and this is in line with best practice employed across the public sector.
8. This would play a pivotal role in fostering this trust by ensuring that AI systems are deployed ethically and responsibly, with the provision of exceptional services for residents at the heart of our approach.
9. We will review and adopt established approaches to ethics to guide the council's approach in this area. Any Members who work on this framework would require training to ensure they are able to apply the framework correctly to guide their decision making.

AI Adoption Models: AI in Productivity tools and AI integrated into line of business applications.

10. It is important to recognise that a distinction between AI introduced into productivity tools such as Microsoft Copilot for Microsoft Office suite and AI integrated into line of business applications, and where there are points of intersection between them, needs to be made to better understand the scope and scale of AI adoption within H&F.
11. AI in productivity tools: The primary focus of productivity tools enhanced by AI such as Microsoft Copilot is on improving individual efficiency. These tools automate routine tasks, enhance document creation, and streamline workflows.
12. H&F have procured a limited number of Microsoft Copilot licences, and strategically assigned these, based on identified use cases and supporting business cases.
13. Microsoft Copilot for Edge has been enabled across the entire organisation, which assists in content generation within our secure environment, thereby mitigating some of the risks associated with public tools such as ChatGPT.

Examples of AI use in productivity tools within H&F

Resident services – Complaints and Microsoft Copilot

14. The Resident services team have adopted Microsoft Copilot for Word to support the complaints management process. This tool leverages Artificial Intelligence (AI) capabilities to help streamline drafting and communications around complaints, offering quick summaries and suggested responses, whilst using reference documents within our secure environment. We are also using it to summarise key performance reports and are exploring its use in the provision of information packs to the Ombudsman.

15. The tactical use of Copilot within the complaints management process creates a limitation on how far these types of AI capabilities can be fully integrated into the complaint management process, particularly those aspects requiring real-time collaboration and more complex analysis and document management.

Next steps and recommendations:

16. It is anticipated that a more strategic approach aligned to the updated AI adoption process within H&F will be adopted. This will involve the development of a robust and scalable complaints management solution, leveraging such advanced AI tools such as Azure AI search to provide deeper insights, collaboration, and data integration to provide a higher level of automation and precision in complaints management.
17. Any further deployment of AI capabilities for complaints management will be reviewed against the AI governance framework.

AI integrated into H&F line of business applications.

18. At this early stage of AI adoption, there will be instances where business units may individually adopt or deploy AI features within their Commercial off-the-shelf applications. The council is working to ensure adoptions or deployments are still be aligned strategically to the aims and objectives of the Resident experience and access programme (REAP). Some line of business applications are covered by the REAP programme directly and the use of AI will be considered here also.
19. AI integrated into line of business applications is designed to transform core business processes, offering capabilities such as predictive analysis, and decision-making support.
20. These applications directly impact operational efficiency, service delivery and customer engagement.
21. The integration of AI within line of business application is still in the early stages within H&F. Our focus has been on assessing and preparing for this development. Products with AI capabilities are currently in the H&F project pipeline and are being reviewed to ensure the necessary guardrails for adoption are in place.
22. Our AI governance framework places us in an excellent position to safely embark on these more complex AI integrations within line of business applications.

Examples of AI and Machine Learning solutions integrated into line of business applications.

H&F Housing team: Voicescape Implementation

23. The Housing team are implementing Voicescape, primarily to leverage the advanced AI and machine learning capabilities that Voicescape offers. The deployment is aimed at enhancing business operations, particularly within rent arrears management and tenant communication processes. Voicescape's AI driven insights allow the team to better analyse tenant payment behaviours, prioritise arrears cases, interventions, and optimise engagement strategies thus increasing efficiency across the business function.
24. The tool has been configured with strict guardrails in line with support and recommendations from the vendor, to ensure that AI is used as a tool for insight. The tool will perform underlying data analysis and generate recommendations using machine learning. The workflow and action is undertaken manually by officers, who exercise their own judgement in applying making final decisions.
25. Voicescape was reviewed against the draft H&F AI governance framework to ensure that the guardrails informing ethical oversight and accountability when deploying such AI and machine learning driven tools are maintained.

Intersection of AI in productivity tools and AI integration into line of business applications integration:

Children's Services – Education, Health, and Care Plans:

26. Children's Services will soon be trialling the AI assisted application for Education, Health, and Care Plan (EHCP) generation developed by Agilisys. The application is designed to provide a scalable and collaborative workspace for generating EHCP plans. This is expected to provide efficiencies and reduce the manual workload typically involved in EHCP plan creation.

Some examples of AI use cases in other local authorities and organisations

27. The use of AI, including both classic and Generative AI, within UK local authorities is expanding, with varying degrees of success driven by the need to improve efficiency, decision making and service delivery.
28. Westminster City Council has implemented a Generative AI solution to handle over 30,000 waste-related issues reported annually by residents. Launched on March 11, 2024, the online chat tool allows residents to report waste, fly-tipping, graffiti, and animal fouling in under a minute. This AI system enhances efficiency, service delivery, and resident satisfaction through a user-centered design approach.
29. Newham Council and the University of East London are working together to develop ways to improve service delivery and address local government challenges. The initial focus is on better predicting housing demand and homelessness trends and streamlining the procurement and allocation of properties.

30. Derby City Council are looking at implementing AI capabilities across multiple services. The council initially partnered with ICS.AI to deploy digital assistants, which use conversational AI to manage customer queries via telephone and the web. These routine queries related to council services such as council tax, parking and benefits reduced staff workloads by automating responses to residents. The council has awarded a £7 million project to expand its AI capabilities, identifying multiple different use cases for the technology. The first phase will target Adult Social Care, Customer Services, and Debt Recovery, using AI to review care packages, streamline customer services, and improve debt recovery processes.

31. Wigan Council is exploring the use of AI technology with senior leadership support and is working in partnership with Agilisys. To date, they have engaged in envisioning sessions to assess how AI could be applied across the council. As part of these efforts, they have started prototyping several use cases and formed a Generative AI Working Group to support the adoption of AI within the council. One of the tools being trialled is the Agilisys Quick Action AI tool, which is used for tasks such as content generation, meeting transcription, and document analysis, with the current focus being on adult social care. In addition, they are reviewing other tools such as Agilisys Nava and Cura classic chatbot AI tools to improve customer experience.

32. We are also reviewing innovative uses of AI just emerging and being tested by several other local authorities. These examples include newly emerging tools which seek to address discrete challenges such as planning transport routes, processing revenues and benefits, summarising case files, or assessing learning in schools.

33. It is important to highlight that our efforts on AI adoption align closely with other local authorities, and indeed with other organisations. In line with them, H&F is approaching adoption with caution including ensuring the necessary guardrails and governance frameworks are in place to support our efforts.

H&F AI Adoption Risks, issues, and challenges

Risks	Impact	likelihood	Mitigation
<p>An ongoing and critical risk identified in the initial briefing paper to POB in March 2024, is the issue of data quality.</p> <p>As we continue our AI adoption journey, the success of AI initiatives will be dependent on the quality, accuracy, and completeness of our data. Poor or incomplete data can greatly affect the efficacy of the AI tool.</p>	High	High	As part of the AI adoption process, a robust data governance framework and initiatives to drive up our data quality should be prioritised and be a major pre-requisite for our AI initiatives.

--	--	--	--

Risks	Impact	likelihood	Mitigation
There is a risk that AI adoption may occur in a tactical and siloed manner, leading to misalignment with the organisation's overall strategic vision.	High	High	To mitigate this risk, AI adoption within business units should be aligned to the AI governance framework and, where relevant, the broader REAP transformation programme. This will ensure consistency, strategic alignment, wider benefits realisation and avoid fragmented efforts across the organisation.
There is a risk that application vendors may introduce AI features and capabilities into their commercial off-the shelf application without the active participation of the business units that use them or the wider organisation. This could result in AI tools that do not effectively support H&F's business outcomes or align with organisational needs.	High	High	As regular updates to applications are rolled out by vendors, any AI features introduced by vendors must be functionally reviewed by the business before adoption, and such capabilities should be disabled by default and only activated once their implications are understood and there are clear benefits to the organisation within the relevant governance frameworks.
Existing contracts for commercial off-the shelf applications may include AI features or capabilities that raise concerns related to costs, ethics, privacy, and efficacy of the product. The organisation may face challenges in determining the impact of these AI features leading to potential issues in cost management, ethical and privacy considerations, information security and assurances around effectiveness.	High	High	The AI Governance framework will be used to review products and address concerns related to cost, privacy, ethics, and efficacy. Additionally, an AI attestation process will be integrated into contract review and procurement processes to ensure AI features meet organisational requirements.

Risks	Impact	likelihood	Mitigation
<p>Climate change 2030</p> <p>The general adoption of AI and Large language Models (LLM) tool and capabilities could hinder the organisation's climate change aspiration, especially considering our carbon neutrality vision for 2030. This risk is compounded by the exponential growth in the adoption of AI and machine learning features within commercial off- the- shelf line of business applications. These AI and Machine learning features are often introduced into these products without direct consultation with customers. This lack of consultation limits our ability to proactively address the carbon impact of such technologies, potentially undermining our sustainability goals.</p>	low	low	<p>Carbon neutrality must be incorporated into product evaluation criteria when selecting and renewing Commercial off-the-shelf applications that include AI and Machine learning capabilities.</p> <p>Adoption of AI tools must be closely aligned with the organisation's strategic vision on carbon neutrality by 2030. Where practicable we must work with Commercial off-the-shelf application vendors to obtain clear attestations on the carbon footprint of their AI and machine learning tools.</p>
<p>Return on Investment (ROI)</p> <p>Demonstrating and attaining a clear return on investment for AI and machine learning capabilities, particularly within the context of productivity tools and AI integrated into line of business applications can be quite challenging and Gartner have written extensively about these challenges.</p> <p>Productivity tools such as Microsoft M365 applications, offer more immediate efficiency gains, through improved workflows, better decision making, and reduced manual workloads. However, feedback from organisations such as Gartner suggests that the realised impact and gains from</p>	High	High	<p>Our approach of ensuring that any use cases are supported by detailed business cases will be carried forward to the redesigned AI adoption process, mandating that all AI use cases include clear benefits realisation, success criteria, and return on investments analysis with metrics on short-, medium- and long-term expected ROI.</p>

these tools do not meet initial expectations as people tend to exaggerate their expected gains by a factor of 2 to 3 times, and any actual productivity gains are split between the individual and their organisation.			
<p>Reputational damage</p> <p>The use of AI and Machine learning tools in decision making carries a significant risk of reputational damage to an organisation such as H&F if these tools generate incorrect outcomes. If such AI-driven decisions lead to wrongful actions, the authority could be severely damaged.</p>	High	High	To forestall such outcomes as described, it is important that AI tools are integrated into local business processes with strong governance and human oversight, ensuring that all recommendations are carefully reviewed before critical actions are taken. The AI governance framework has been designed to aid and reinforce such essential and core principles in the future use of AI and machine learning tools.

LEGAL IMPLICATIONS

Use of AI must comply with the UK GDPR and Information Commissioner Office guidance. Each contract where AI is a component part should deal with

- risks and consequences, which party is best placed to manage risks and bear liability, what should be excluded/limited.
- KPIs/remedies for failure
- Clear governance
- Warranties
- Transparency/accountability/remedial actions when things go wrong.

In addition, particular care should be taken in respect of automated decision making affecting individual rights. Article 22 of the GDPR provides that ‘The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her’. Automated decision making in these circumstances is unlikely to be lawful unless it has been specifically authorised by statute or consented to (see Master of Rolls speech 9th October 2024 [Speech by the Master of the Rolls: AI and the GDPR - Courts and Tribunals Judiciary](#)).

Angela Hogan, Chief Solicitor (Contracts and Procurement) 16th October 2024

APPENDIX A

H&F DRAFT AI GOVERNANCE FRAMEWORK.

BUSINESS CASE AND RISK ASSESSMENT FOR GENERATIVE AI AND TRADITIONAL AI

BACKGROUND

It is crucial for all staff to be aware of the governance processes and steps that must be completed when submitting a business case for an AI solution. These procedures are designed to maintain standards of regulatory compliance, security, and risk management, as well as important supporting ethical decisions on the adoption of technology, ensuring that AI technologies are deployed effectively and responsibly within the council. They build on the information governance and cyber security approaches already in place.

When completing your business case, you are obliged to complete the risk assessment section fully and following these key processes:

- Project Concept Document (PCD) Phase
- Technical Design Authority (TDA) review questions
- Non-Functional Requirements (NFRs)
- Completion of DPIA where personal data is processed.
- Completion of SSQ where external supplier will be involved.

PURPOSE OF THIS DOCUMENT

Due to the advancement and proliferation of AI technologies available, the council is developing a governance framework that supports the responsible and risk-based adoption of these technologies. This framework ensures that the adoption of AI within the council is subjected to a rigorous and consistent risk assessment process. All staff are required to complete the AI business case and risk assessment, following the relevant governance processes, for both In-house developed solutions such as low code platforms as well as commercial off the shelf (COS) products and services.

INTRODUCTION

Generative AI

Generative AI is a branch of artificial intelligence that can create new content or data based on existing examples or rules. Examples of generative AI include natural language generation, image synthesis, music composition, and more. Generative AI has the potential to enhance creativity, productivity, and innovation.

Traditional AI

Traditional AI uses symbols and rules to show and change information and uses search methods to find answers in a fixed set of options. An example of traditional AI is a chess program that uses predefined rules and strategies to choose the best move from a finite set of moves. The chess program represents the board and the pieces as symbols and applies logical rules to manipulate them and evaluate the outcomes. The chess program also uses search methods to explore the game tree and find the optimal move. Unlike Generative AI, Traditional AI program does not learn from its own experience or adapt to new situations but follows the fixed rules and algorithms that are programmed into it.

The use of AI, traditional or generative

Whether implementing Traditional AI or Generative AI, both technologies pose some challenges and risks that need to be carefully considered and managed before adoption within H&F. It is important to have a clear and comprehensive business case and risk assessment for the use of AI so H&F has considered data protection, information security, ethics and how the solution will help H&F and our residents.

This document provides a template for creating a business case and a risk assessment for the use of AI within H&F. The template is based on the following steps:

- Define the problem or opportunity that AI can address.
- Identify the objectives and benefits of using AI.
- Analyse the alternatives and costs of using AI.
- Evaluate the risks and mitigation strategies of using AI.

BUSINESS CASE

The completion of this business case and risk assessment is a requirement of the Project Concept Document Phase as part of the H&F Digital Services Project Delivery Framework. Your AI business case should answer the following questions:

DEFINE THE PROBLEM OR OPPORTUNITY

- What is the problem or opportunity that the project or initiative aims to address?

Describe the current situation and the gap or need that the project or initiative aims to address. Explain why the problem or opportunity is important and relevant to the project or organisation.

IDENTIFY THE OBJECTIVES & BENEFITS

- What are the objectives and benefits of the project or initiative?

Define the specific and measurable goals and outcomes that the project or initiative expects to achieve. Quantify the benefits and value that the project or initiative will deliver to the project or organisation.

ANALYSE THE ALTERNATIVES & COSTS

- What are the alternatives and costs of the project or initiative?

Identify and compare the different options or solutions for addressing the problem or opportunity. Include the use of AI as one of the options. Estimate the costs and resources required for each option, including the initial and ongoing expenses.

e.g. additional staff to implement the solution or write prompts.
e.g. paying for new licenses for using AI software
e.g. costs associated with training staff to use the new solution.
e.g. costs associated with organisational restructuring.

EVALUATE THE RISKS AND MITIGATIONS

- Where you have identified a need for AI, you have liaised with your [Strategic Relationship Manager](#) (SRM)
- You have engaged with and followed the Project Concept Document Phase Process
- You have fully completed the AI risk assessment template.
- Where you have identified whether the proposed use of AI will involve the processing of personal data and if so, you have completed a [Data Protection Impact Assessment](#).
- Where you have identified that the use of AI will involve a third-party supplier processing H&F data, you have submitted a completed [supplier security questionnaire](#)
- If you are procuring a new solution, you have incorporated the H&F Non-Functional Requirements (NFRs) into the requirements for your solution. You can obtain these from your SRM.
- You have submitted your proposal for AI to the Technical Design Authority for technical review.

- You have obtained comments from the Senior Information Risk Owner, and Chief Digital Officer.
- You have obtained comments on the ethical implications of your proposed use of AI from People & Talent
- You have scheduled a meeting for your proposal to be reviewed at the Digital Ethics Board (if your use of AI is resident facing/likely to have an impact on residents).

AI RISK ASSESSMENT TEMPLATE

DATA PROTECTION RISK:

Non-compliance with the UK GDPR and DPA 2018 can pose a serious risk to the data subjects whose personal data is processed by AI, as well as to the data controller (H&F) who is responsible for ensuring the lawfulness, fairness, and transparency of the processing. The consequences of non-compliance with the UK GDPR and DPA 2018 can include:

- Harm to data subjects (individuals whose data is processed by the AI model)
- Complaints, claims, or litigation from the data subjects or other stakeholders, which may damage the reputation, trust, of the data controller and the use of AI.
- Enforcement actions, fines, or sanctions from the Information Commissioner's Office (ICO)

Answer the questions below to determine whether you need to complete a [Data Protection Impact Assessment \(DPIA\)](#).

If you answer yes to any of these questions you must complete a DPIA and submit it to the information Management Team (informationmanagement@lbhf.gov.uk)

#	Question – will the project involve the processing of personal data. To determine please answer these screening questions.	Yes	No
1	Will the project involve the collection of new information about individuals?		
2	Will the project compel individuals to provide information about themselves?		
3	Will information about individuals be disclosed to organisations or people who have not previously had routine access to the information?		
4	Are you using information about individuals for a purpose it is not currently used for, or in a way it is not currently used?		
5	Does the project involve you using new technology which might be perceived as being privacy intrusive? For example, Generative AI, Traditional AI, CCTV, the use of biometrics or facial recognition.		
6	Will the project result in you making automated decisions about individuals?		
7	Will the project result in you making decisions or taking action against individuals in ways which can have a significant impact on them?		
8	Is the information about individuals of a kind particularly likely to raise privacy concerns or expectations? For example, health records, criminal records, or other information that people would consider to be particularly private.		
9	Will the project require you to contact individuals in ways which they may find intrusive?		

THIRD PARTY SUPPLIER SECURITY RISK:

An assessment of the supplier's security can help ensure compliance with applicable standards and regulations, and establish appropriate controls are in place to safeguard the council's data and assets. By doing an information security risk assessment using H&Fs Supplier Security Questionnaire (SSQ) on new suppliers,

the council can reduce the likelihood and impact of any cyber incidents that may compromise the confidentiality, integrity, or availability of the information or services provided by the supplier.

Are you using a third-party solution to perform any operation (including storage) on H&F data? If so, you will need to complete a Supplier Security Questionnaire (SSQ) and submit it to the IM team (informationmanagement@lbhf.gov.uk)

[IM - Information Security: Template - Supplier Security Questionnaire](#)

Additionally, due to the absence of standardised certifications or kite marks that differentiate between well designed or poorly designed AI, we will undertake a more nuanced view when evaluating these systems, focusing on factors such as reputation of the AI provider, track record of developing AI tools, any implied trust based on previous experiences.

ACURACY RISK:

AI may produce incorrect, misleading, or fabricated content or data that can affect the quality, reliability, or credibility of the information. This can have negative consequences for the users, consumers, or decision-makers who rely on the generated output, such as misinformation, deception, fraud, or legal liability. This risk is associated with generative AI but is also applicable to traditional AI.

Describe below how you will mitigate this risk for your proposed use of AI:

e.g. review, verify, correct, or reject the AI outputs, and to report any errors, inconsistencies, or inaccuracies to the system developer or provider.

e.g. compare AI outputs with human-generated or verified outputs.

e.g. ensure that the data used to produce the AI outputs has a high level of accuracy and quality, and where it does not ensure that this is made clear in your findings.

e.g. undertake a data quality and cleansing exercise on the data to be used to generate AI outputs before the system or process is live.

TRANSPARENCY RISK

AI may produce outputs that are not easily explainable, interpretable, or traceable by the users, consumers, or decision-makers who use them. This can affect the trust, accountability, or responsibility of the technology and its outcomes, especially when it involves sensitive or high-stakes domains, such as healthcare, education, or justice. Therefore, it is important to ensure that the AI system provides sufficient transparency about its data sources, methods, limitations, and uncertainties, and that the users can access, understand, and challenge the generated output if needed.

Describe below how you will mitigate this risk for your proposed use of AI:

e.g. Tell those whose data is processed by AI where the data comes from, how the system works, what are the limits and uncertainties, and make this information easy to find and understand.

e.g. Let the users check or question the outputs of AI if they are not sure or have worries. (for example, include a message in automated communications to advise how they can contact H&F to query the outputs)

e.g. Make sure you have ways for users to give feedback, or complain, and let them report or flag any problems or errors with the output.

e.g. Ensure users know that the outputs are from generative AI by using labels, warnings, or different colours or fonts.

e.g. Add to privacy notices/fair processing notices that decisions have been made using automated decision making and an explanation of what data and logic is used to make the decision.

BIAS RISK:

AI may produce outputs that reflect or amplify the biases, stereotypes, or prejudices that exist in the data or the model. This can result in unfair, discriminatory, or harmful outcomes for certain groups of people or individuals, such as excluding them from opportunities, misrepresenting them, or affecting their dignity or well-being. This risk is associated with generative AI but is also applicable to traditional AI.

Therefore, it is important to ensure that the outputs are monitored and evaluated for potential bias, and that the users are aware of the limitations and assumptions of the system.

Describe below how you will mitigate this risk for your proposed use of AI:

e.g. Incorporate human oversight and intervention mechanisms to review and validate the outputs of the AI system, and allow the users to modify, reject, or report the outputs if they find them inappropriate, inaccurate, or biased.

HALLUCINATIONS RISK:

Generative AI can make content up that is not based on the data. This can make the outputs wrong, misleading, or nonsensical. The risk with hallucinations is that these outputs may be mistaken for accurate information leading to incorrect decisions, or presumptions.

e.g. use other sources and references to compare and confirm the outputs of generative AI and find and fix any mistakes or hallucinations.

e.g. report any issues with the system such as hallucinations to the system developer

ETHICAL RISK:

Generative AI can be used unethically to produce false or misleading information that can harm the reputation, privacy, or security of individuals, organisations, or society. For example, generative AI can produce deepfakes, which are realistic but fabricated images or videos of people, which can be used for impersonation, fraud, or defamation. There is also a risk of violating the intellectual property rights of the original creators or sources of the data used by the generative AI system, such as through plagiarism. This risk is associated with generative AI but is also applicable in some cases to traditional AI, for example where you are making significant decisions about individuals using AI models.

Describe below how you will mitigate this risk for your proposed use of AI:

e.g. Wherever requesting consent from individuals to use AI make sure you respect their preferences about how their data is shared.

e.g. use anonymised data rather than personal data to protect the privacy of individuals.

e.g. review all outputs to ensure that personal data is not unintentionally captured in the outputs.

e.g. avoid relying on the outputs of AI for decisions about individuals – all decisions should be made with a human in the loop (a human being making the final decision)

e.g. keep track and evaluate the outcomes of your deployment of AI noting the possible ethical implications of your usage.

e.g. Raise ethical concerns with the H&F AI Ethics & Governance Board.

Technical Design Authority

- **Date reviewed:**
- **Approval Decision (Yes/No/ Amendments needed):**
- **Comments:**

Senior Information Risk Owner (SIRO) Comments

Comments should focus on whether the proposed solution will align with H&Fs Digital and Information Strategy, meet digital accessibility requirements, align with H&Fs technical infrastructure, and meet H&Fs data protection and information security compliance obligations.

People & Talent

Comments should consider the response to the ethical risk section when determining whether the risks identified are proportionate to the impact on members of staff when balanced against the costs/savings and benefits.

AI Governance Process

