Science Council Report of Project 'Artificial Intelligence Applications in Food Safety and Authenticity'

Science Council Report of Project 'Artificial Intelligence Applications in Food Safety and Authenticity'

Food Standards Agency Science Council Report of Project ‘Artificial Intelligence Applications in Food Safety and Authenticity’

In this guide

In this guide

1. Food Standards Agency Science Council Report of Project ‘Artificial Intelligence Applications in Food Safety and Authenticity’
2. Authors, Acknowledgements and Declarations of Interest
3. Executive Summary
4. Recommendations to the FSA
5. Introduction
6. Methodology
7. Results and Discussion
8. Conclusions
9. Appendix A: Workshop Case Study Briefing Document
10. Appendix B: Workshop Participant List
11. Appendix C: Workshop Case Study Responses
12. References

https://doi.org/10.46756/sci.fsa.yne470

Science Council

Report of Project

‘Artificial Intelligence Applications in Food Safety and Authenticity’

Published: 28 October 2025

Science Council Report of Project 'Artificial Intelligence Applications in Food Safety and Authenticity'

Authors, Acknowledgements and Declarations of Interest

In this guide

In this guide

1. Food Standards Agency Science Council Report of Project ‘Artificial Intelligence Applications in Food Safety and Authenticity’
2. Authors, Acknowledgements and Declarations of Interest
3. Executive Summary
4. Recommendations to the FSA
5. Introduction
6. Methodology
7. Results and Discussion
8. Conclusions
9. Appendix A: Workshop Case Study Briefing Document
10. Appendix B: Workshop Participant List
11. Appendix C: Workshop Case Study Responses
12. References

Authorship

This report of the FSA Science Council project on Artificial Intelligence Applications in Food Safety and Authenticity was prepared and reviewed for publication by the Science Council on 23 September 2025.

Science Council

- Project Chair: Professor Simon Pearson

- Project Leads: Professor Peter Gregory, Mrs Claire Nicholson, Professor John O'Brien (Science Council Chair)

- Project Team: Professor Peter Borriello CB, Professor Emily Burton, Professor Steven Cummins, Jacqueline Healing, Professor Tom Oliver, Professor Richard Smith, Professor Michael Tildesley, Professor Jonathan Wastling

Acknowledgements

The authors thank the Food Standards Agency (FSA) Science Council and Secretariat for their guidance and support in the preparation of this report. We are also grateful to the workshop participants and external experts who contributed to the case studies on AI applications in food safety and assurance. Any errors or omissions remain the responsibility of the authors.

Declarations of Interest

In line with FSA Guidance on managing interests of its scientific advisers, the interests of the Science Council members were assessed to identify any potential conflicts with the work of this Group.

Note that a full register of interests for all Science Council members can be found at Science Council Register of Interests | Science Council.

Professor Simon Pearson specifically declared that he is a Director of Fruitcast.ai, an AI company specializing in crop forecasting, and of Agaricus Robotics, a robotics company that deploys AI in control systems for mushroom harvesting. Professor Peter Gregory, Mrs. Claire Nicholson and Professor John O'Brien declared no relevant interests in relation to this review.

The Executive considered the potential interests of Science Council members, and they were not considered prohibitive for their involvement in this review by the Executive and would not require specific measures to manage them.

Disclaimer

This report has been prepared by members of the Science Council of the Food Standards Agency. It does not necessarily represent the views of the Food Standards Agency or its Board. The content is based on the independent judgement of the Council members and the evidence available at the time of writing.

Science Council Report of Project 'Artificial Intelligence Applications in Food Safety and Authenticity'

Executive Summary

In this guide

In this guide

1. Food Standards Agency Science Council Report of Project ‘Artificial Intelligence Applications in Food Safety and Authenticity’
2. Authors, Acknowledgements and Declarations of Interest
3. Executive Summary
4. Recommendations to the FSA
5. Introduction
6. Methodology
7. Results and Discussion
8. Conclusions
9. Appendix A: Workshop Case Study Briefing Document
10. Appendix B: Workshop Participant List
11. Appendix C: Workshop Case Study Responses
12. References

Artificial Intelligence (AI) technologies are advancing rapidly with significant potential to transform how food is produced, managed and regulated. In the food system, AI offers the prospect of more efficient, predictive and responsive assurance processes, ranging from real-time detection of hazards to automated documentation checks and predictive modelling of risks. At the same time, the adoption of AI raises important questions about accountability, transparency and trust, particularly where automated systems interact with actions to comply with the Food Safety Act (1990) and the due diligence defence relied upon by Food Business Operators (FBOs).

The Food Standards Agency (FSA) has a responsibility to ensure that innovation does not undermine consumer protection, regulatory oversight, or public confidence. This Science Council project was established to explore the likely applications of AI in food safety and assurance; identify the benefits and risks and consider implications for the FSA’s role as a regulator. The study drew on academic and policy evidence with key insights from a June 2025 workshop attended by food businesses, regulators, assurance providers, academics and technology developers. Four case studies were used to anchor discussions in realistic scenarios where AI (all forms from machine learning applications to generative and emerging AI systems) is likely to be applied now and in the near future:

- AI-driven safety and regulatory compliance evaluation for manufactured foods
- AI-supported data pack generation for third-party certification and assurance
- AI-assisted detection of infections and other pre/post-mortem pathologies in UK abattoirs
- AI-powered document inspection at UK ports of entry

These case studies, and post-workshop discussions amongst the project team, wider science council and FSA staff, highlighted both opportunities and challenges. AI could enable faster detection of hazards, more consistent and scaled inspections, wider surveillance across supply chains and real-time data analytics to help target interventions. It could reduce reliance on sampling or retrospective checks and free inspectors or auditors to focus on higher-value tasks. However, workshop participants also identified risks: AI systems may embed bias or drift if not carefully validated; they can generate outputs that are difficult to explain or reproduce; and in poorly managed businesses, they could conceal weaknesses behind apparently robust documentation. Across all scenarios, the need for human oversight, clear accountability, explainability of decisions and robust validation of AI tools was consistently emphasised. Especially in these early stages of AI deployment, when both industry and regulators are still learning how AI systems can be safely applied, vigilance is essential; over time, experience will help clarify the contexts in which AI delivers most value and the necessary safeguards.

For the FSA, the implications are clear. AI has the potential to strengthen assurance processes, but only if deployed within strong governance frameworks and supported by clear guidance. We recommend the Agency should clarify best practice for the integration of AI within FBO accountability; continue its promotion of data standards and sharing; provide guidance for FBOs on responsible use; and work with industry, standards bodies, and other regulators to support codes of practice and validation mechanisms. At the same time, it must remain alert to the risks of hype and over-reliance, ensuring that AI enhances, rather than displaces, the human accountability that underpins food law.

Due to the rapid emergence of AI technologies, and notwithstanding the present limited case study evidence from scaled industrial use, there was broad consensus that existing UK food safety regulations are sufficiently robust to encompass the use of currently known AI systems in the food system. This study does not therefore call for immediate changes to legislation. However, the FSA will need to continually monitor developments in AI, assess their impacts on assurance processes and remain prepared to act if gaps emerge. Future regulatory attention may be required in areas such as validation standards, data governance, or liability frameworks should AI adoption accelerate, or if new classes of tools present novel risks.

The following recommendations aim to provide the FSA with practical steps to support safe and responsible AI adoption, ensuring that innovation contributes to a more predictive, preventative, and trusted food safety system.

Recommendations to the FSA

In this guide

In this guide

1. Food Standards Agency Science Council Report of Project ‘Artificial Intelligence Applications in Food Safety and Authenticity’
2. Authors, Acknowledgements and Declarations of Interest
3. Executive Summary
4. Recommendations to the FSA
5. Introduction
6. Methodology
7. Results and Discussion
8. Conclusions
9. Appendix A: Workshop Case Study Briefing Document
10. Appendix B: Workshop Participant List
11. Appendix C: Workshop Case Study Responses
12. References

1. Publish Guidance on Responsible Use of AI to Assure Food Safety and Regulatory Compliance

The FSA should provide guidance for Food Business Operators (FBOs) making clear that AI must act only as a decision-support tool, with humans retaining accountability. AI outputs should remain transparent and explainable to underpin regulatory compliance and ensure safety-critical decisions are not delegated to AI agents of dependent systems. Food businesses should be aware of limitations and work with suppliers to ensure tools are fit for purpose.

2. Establish Ongoing Monitoring of AI Systems and Potential Impacts

The FSA’s Strategic Insight Team should regularly monitor AI adoption by Food Business Operators and its impacts across the food system, providing early warning of risks and ensuring regulatory responses remain proactive.

3. Promote Data Assurance, Validation, and Standards Alignment

The FSA should continue its promotion of data quality, provenance and standards (including data and cybersecurity) to help realise fair, auditable AI and, including the promotion of measures to ensure SMEs and smaller FBOs are not disadvantaged.

4. Support the Development of Standards and an Industry-Led Code of Practice for Assuring AI in Food Safety

The FSA should promote the development of test and validation standards for AI systems and a Code of Practice to food businesses, technology suppliers, standards bodies, potentially acting as a convenor to ensure alignment with regulatory expectations and wider assurance frameworks.

5. Engage with Broader Regulatory and Policy Perspectives

The FSA should collaborate with other UK and international regulatory authorities and government agencies to align governance, ethics, and auditability standards, drawing lessons from industries with advanced AI regulation.

6. Commission Research on Human Behaviour and Interaction with AI in Food Safety Contexts

The FSA should advocate for behavioural research that studies how FBO workers and directors engage with AI, addressing issues such as overreliance, trust, and cognitive bias to inform better training, guidance and governance.

Science Council Report of Project 'Artificial Intelligence Applications in Food Safety and Authenticity'

Introduction

In this guide

In this guide

1. Food Standards Agency Science Council Report of Project ‘Artificial Intelligence Applications in Food Safety and Authenticity’
2. Authors, Acknowledgements and Declarations of Interest
3. Executive Summary
4. Recommendations to the FSA
5. Introduction
6. Methodology
7. Results and Discussion
8. Conclusions
9. Appendix A: Workshop Case Study Briefing Document
10. Appendix B: Workshop Participant List

11. Appendix C: Workshop Case Study Responses
12. References

Artificial Intelligence (AI) technologies are rapidly transforming both society and industry, reshaping how food is produced, selected and consumed. While the pace of technological development and adoption presents major commercial opportunities, it raises important questions about the safe and responsible use of AI across the food system. The objectives of this FSA Science Council report are to anticipate likely impacts of known and emerging AI systems and to assess potential implications for food safety and assurance. It considers how AI could affect critical food safety functions and explores perspectives on the standards required for AI function. This report represents the FSA's first formal examination of AI in the food system. However, we acknowledge that the risks and perceptions of this diverse and rapidly evolving technology will become clearer as adoption increases and its intended, and unintended, consequences are known.

Modern AI represents a family of technologies that includes machine learning, computer vision, robotics, natural language processing, and large language model typologies (LLMs), each offering distinct capabilities for tasks such as detecting foodborne risks, automating visual inspections, interpreting regulatory documents, translating multilingual records, and extracting insights from complex or unstructured data. Applications may engage a single AI function or multiple interconnected technologies, for example, combining computer vision for image recognition, machine learning for pattern detection, and large language models (LLMs) for interpreting documentation or generating decision support. The evolution of these technologies has occurred at astonishing pace. Key breakthroughs include AlexNet (Krizhevsky et al. 2012), where deep convolutional neural networks dramatically improved image recognition accuracy over classic computer vision techniques, and large language models (LLMs), highlighted with the launch of ChatGPT as recently as November 2022 (OpenAI, 2022), which brought generative AI into widespread public view and commercial use.

AI adoption across the UK food system is accelerating, particularly in food manufacturing, logistics, and primary production, where technologies such as machine learning, computer vision, and robotics offer productivity gains. In manufacturing, AI is likely to improve production efficiency, safety compliance and quality control; potentially detecting non-conformances more reliably and at greater scale than manual checks, while also reducing labour costs and waste. Newer approaches, such as imitation (Li et al., 2025) and reinforcement learning, will enable robots to mimic complex and dextrous human behaviours found in harvesting, handling, and inspection tasks that are beyond current state-of-the-art machines. These advances support not only greater automation but also improved responsiveness to changing supply chain conditions and consumer demands. As AI systems become more accessible and interoperable, they are expected to underpin a shift toward more adaptive, data-driven decision-making across the entire food system.

AI has the potential to transform food safety by shifting assurance processes from largely reactive responses to more proactive, predictive and real-time management. Emerging applications include predictive analytics to anticipate pathogen risks (Benefo et al., 2022), computer vision systems that can continuously monitor processing environments for hygiene and non-conformances (Zhao et al., 2025), and digital twins that model facility operations to optimise preventive controls (Pennells et al., 2025). AI has demonstrated potential to improve both the chemical and microbiological safety of food. Machine learning has shown the potential to improve source attribution in foodborne outbreaks when combined with whole genome sequencing (Munck et al., 2020). A recent review by Kabir et al. (2025) suggested that machine learning applied to hyperspectral imaging data had potential to classify grains and nuts according to mycotoxin contamination. A significant body of research is already available in this area using many different ML algorithms. AI can also extend surveillance beyond the factory floor, using natural language processing to analyse consumer complaints or social media signals, and machine learning to integrate disparate datasets into early-warning systems for contamination or fraud (Tao et al., 2021). AI could also be used to both generate but also detect fraudulent activity (e.g. fake documents, fake certificates, fake labels etc); positively it could enable regulators and food businesses to respond faster, and reduce reliance on sampling or retrospective testing.

AI offers opportunities to enhance consistency and scale. Unlike human inspectors who must work within time and resource limits, AI systems can continuously scan large volumes of multimodal data (images, text, sensory data etc) across production lines or even supply chains, potentially identifying trends and anomalies invisible to individual auditors. By supporting human decision-making with richer evidence, AI could reduce variability between inspectors, increase sampling rates whilst enabling more transparent traceability from farm to fork. Whilst many consider these applications remain unevenly developed, they highlight the direction of travel: food safety may become more predictive, more integrated, and more responsive as AI tools mature. In short, the technology has the capacity to enhance both the efficiency and resilience of assurance systems, provided it is deployed with the appropriate safeguards. However, the translation of these capabilities into real-world settings must be handled carefully. Depending on how AI is introduced, it could lead to significant changes in worker roles or perceived redundancy of certain tasks, raising serious concerns with jobs and needs for reskilling.

At the same time, the integration of AI into food safety and assurance raises fundamental questions about accountability, explainability, and trust. The FSA operates within a robust legal and regulatory framework, underpinned by the Food Safety Act (1990), which places ultimate responsibility for food safety on the human decisions made by employees and directors of Food Business Operators (FBOs). This accountability cannot be transferred to an algorithm. As the Law Commission (2025) has recently highlighted, autonomous and adaptive AI systems “do not currently have separate legal personality… [and] could lead to ‘liability gaps’, where no natural or legal person is liable for the harms caused” (Law Commission, 2025). This concern is highly relevant to the food system, where unexplained or unverifiable AI outputs could undermine both consumer protection and the due diligence defence relied upon by FBOs.

These considerations mean that the deployment of AI in food safety must be accompanied by clear governance, transparency, and human oversight. The challenge for the FSA is to balance innovation and efficiency with regulatory assurance, ensuring that AI augments rather than replaces the human accountability that underpins food law.

Recent research has emphasised that ethical considerations are inseparable from the deployment of AI in food systems. Manning et al. (2022) argue that adoption of AI will only be trusted if it is grounded in a shared vocabulary of ethical principles that stakeholders across the supply chain can understand and apply. Their review identifies seven interlinked aspects, transparency, traceability, explainability, interpretability, accessibility, accountability and responsibility, as central to embedding AI in food governance. Importantly, they highlight that failure to differentiate or operationalise these aspects risks creating barriers to adoption, undermining trust and amplifying bias. For regulators such as the FSA, these findings underline that the introduction of AI in food assurance is not simply a technical question but also a socio-ethical challenge: AI must be explainable, accountable and accessible in ways that align with existing food safety responsibilities if it is to support, rather than erode, consumer confidence (Manning et al., 2022).

Complementing this ethical perspective, Qian et al. (2023) highlight the breadth of AI applications emerging in food safety. They emphasise that adoption remains limited compared to other areas of the agri-food system, constrained by fragmented data sharing, privacy and commercial sensitivity concerns, lack of standardisation, and the absence of clear legal frameworks. Many systems remain at the research stage, often product- or pathogen-specific, with limited scalability into operational practice. Addressing these barriers will require investment by businesses in digital infrastructure, harmonisation of data standards, and frameworks that safeguard both privacy and regulatory integrity.

Taken together, these studies reinforce a common conclusion. AI will be most effective in food safety as a decision-support system operating under human oversight, embedded in strong ethical, legal and governance structures, rather than as a replacement for human accountability. The risks of AI in the food system are not confined to technical performance; they extend to how accountability is assigned, how outputs are explained, and how governance mechanisms maintain oversight.

While AI offers real opportunities to enhance food safety, there is a parallel risk that overstatement or hype could undermine trust in the technology. If inflated claims are allowed to dominate, they risk damaging the reputation of AI before its genuine benefits can be realised. There have been many published cautions about the importance of separating hype and exaggerated claims about AI from the reality (Huckins, 2025 Shoham, 2025). Commentators have shared cases where AI has resulted in unexpected outcomes. To date, most incidents have been relatively small, but they argue for an open but cautious approach. The advent of so called “agentic AI” appears to be at the high end of the hype curve currently and it has been pointed out that, as yet, there is no shared definition of an “agent” in AI (Shoham, 2025). However, AI agents are characterized by combining the power of AI (e.g. LLM) with the specificity of a task (e.g. booking a ticket). In a food safety context, deployment of similar tools would need robust guardrails and close supervision. It is not currently possible to foresee where such a tool could be deployed in food assurance.

Whilst it is becoming clear that AI has transformational power within the food system, its safe and effective adoption will depend on addressing a set of persistent challenges. Questions of data quality, interoperability, transparency, accountability, and legal liability remain central, and the balance between innovation and assurance will be critical. For the FSA, this means considering not only how AI might strengthen food safety controls, but also how its deployment might create new risks, shift responsibilities, or alter the operation of due diligence defences under existing law.

To explore these issues in depth, this report draws on a series of case studies examining the deployment of AI by FBOs across diverse food safety contexts, including product risk assessments, certification and assurance audits, pathology detection in abattoirs, and import documentation checks at ports of entry. Whilst the focus of the report was on FBO application of AI, there is no doubt some of the tools deployed are likely to assist regulators. The case studies, while not exhaustive, provided insights into both the opportunities and risks of AI, illustrating where the technology might augment food safety processes, where it might complicate them, and what governance principles will be needed to ensure safe, fair and trusted adoption across the UK food system.

Science Council Report of Project 'Artificial Intelligence Applications in Food Safety and Authenticity'

Methodology

In this guide

In this guide

1. Food Standards Agency Science Council Report of Project ‘Artificial Intelligence Applications in Food Safety and Authenticity’
2. Authors, Acknowledgements and Declarations of Interest
3. Executive Summary
4. Recommendations to the FSA
5. Introduction
6. Methodology
7. Results and Discussion
8. Conclusions
9. Appendix A: Workshop Case Study Briefing Document
10. Appendix B: Workshop Participant List
11. Appendix C: Workshop Case Study Responses
12. References

The purpose of this study was to examine how artificial intelligence (AI) might be applied in food safety and assurance, to identify the opportunities it offers, the risks it presents, and to assess the implications for the FSA in its role as regulator. The overall objective was to generate evidence-based recommendations to guide the FSA in supporting the safe and responsible adoption of AI across the UK food system.

A central component of the methodology was a full-day workshop held in London on 9 June 2025, attended by 43 participants drawn from food businesses, regulators, assurance providers, academics and technology developers. To structure the discussion, four case studies were prepared in advance, each presenting hypothetical uses of AI in a realistic food industry setting (see Appendix A). The case studies were deliberately framed to address different parts of the food chain, each raising distinct assurance challenges, and each requiring a multitude of diverse AI systems to tackle complex problems. They were selected because they represent both the diversity of the food system and situations where AI deployment is likely to become a reality in the near future. The four scenarios covered:

- AI-driven safety and regulatory compliance evaluation for manufactured foods
- AI-supported data pack generation for third-party certification and assurance
- AI-assisted detection of infections and other pre/post-mortem pathologies in UK abattoirs
- AI-powered document inspection at UK ports of entry

Participants received a briefing pack in advance, which set out the purpose of the study and the issues for consideration (Appendix A), while workshop facilitators were provided with a supplementary briefing document to ensure consistency in the conduct of breakout sessions. Together, these materials provided a shared frame of reference and ensured that discussions were anchored in practical challenges directly relevant to the FSA’s statutory remit.

Participants were assigned to breakout groups, each facilitated by a senior expert and supported by a notetaker. Sessions were conducted under the Chatham House rule to encourage open discussion. Two rounds of breakout discussions, each lasting 75 minutes, allowed participants to participate in two different case studies thereby providing a broad range of perspectives. Each breakout group concluded with the production of a short, summary report that was presented in the final plenary session. These reports enabled findings to be compared across groups and key themes to be determined. To complement the workshop outputs, participants were also invited to submit written reflections after the event, identifying what they regarded as the three most important issues for the FSA to consider.

Analysis of Results

The workshop generated a large volume of qualitative material, including detailed notetaker records from each breakout group, facilitator summaries presented in plenary, and post-event written reflections submitted by participants. These outputs were collated and reviewed to identify both case-specific insights and cross-cutting themes (see Appendix B). Analysis proceeded in two stages. First, the outputs for each case study were organised around the structured questions set out in the briefing materials, ensuring that the findings reflected the issues most relevant to the FSA's remit. Second, themes that cut across case studies were identified, such as the need for transparency, human oversight, validation of training data, and mechanisms to manage bias or drift. These themes informed the synthesis presented later in this report and underpin the recommendations to the FSA.

Evidence from the workshop was then synthesised with relevant literature and policy analysis to draw out cross-cutting issues and to situate the findings within the wider ethical, technical, and legal context. Key areas of analysis included the role of AI as decision support versus autonomous decision-making, the requirements for transparency, explainability and traceability, the challenges of data quality and standardisation, and implications for accountability and legal liability. The approach ensured that the report reflects both expert evidence and stakeholder perspectives, providing a balanced assessment of how AI could shape food safety and assurance in the years ahead.

The case study findings documented in Appendix B present the results of this analysis. Each section begins with the key questions posed, followed by a summary of discussion, supported where appropriate by anonymised quotations. This structure allows both the breadth of perspectives and the areas of convergence or divergence to be captured, providing a balanced account of how AI might realistically shape food assurance.

Results and Discussion

In this guide

In this guide

1. Food Standards Agency Science Council Report of Project ‘Artificial Intelligence Applications in Food Safety and Authenticity’
2. Authors, Acknowledgements and Declarations of Interest
3. Executive Summary
4. Recommendations to the FSA
5. Introduction
6. Methodology
7. Results and Discussion
8. Conclusions
9. Appendix A: Workshop Case Study Briefing Document
10. Appendix B: Workshop Participant List
11. Appendix C: Workshop Case Study Responses
12. References

This section is structured around the major themes that arose from the workshop and subsequent deliberations by the Project Team. Each theme leads, in turn, to a recommendation.

Adoption of AI tools by food businesses

Guidance to FBOs should be issued by the FSA to explain the underlying principles for the use of AI technologies within food safety and assurance processes, including minimum performance expectations, legal responsibilities, risk monitoring, documentation requirements, and clear criteria for human oversight. Ultimately all safety critical decisions should be made by humans, be explainable and traceable.

The case studies reinforced this principle across different contexts. In Case Study 1 (risk assessment of manufactured foods), participants warned that AI could lull businesses into a false sense of security, creating outputs that look convincing but are unverified. Case Study 2 (third-party certification) suggested how AI might process incomplete records, underlining the need for human judgement to challenge results and assess context. In Case Study 3 (abattoirs), the importance of human oversight was highlighted in safeguarding against drift and ensuring ambiguous/unusual cases were resolved by inspectors.

Thus, accountability and business process ownership may be even more important with the advent of AI where there may be a danger of human workers leaving tasks to AI tools without adequate critical supervision. Safety assurance process owners should be human. The individuals concerned should be explicitly identified and competent for that role. While tasks can be assigned to AI algorithms, these should be under human supervision applying appropriate validation and documentation of the process and routine verification to assure that the AI tools perform correctly.

AI systems may support decision-making through data analysis, pattern recognition, or anomaly detection, but must not replace human judgement in safety-critical contexts such as Hazard and Critical Control Point (HACCP) decision points or regulatory inspections. Best practice will include transparent system logs that distinguish between AI-generated outputs and human decisions. This ensures traceability, supports due diligence defences under the Food Safety Act 1990, and maintains public and regulatory trust in AI-augmented assurance systems.

The business relationship between FBOs and AI technology suppliers needs to be carefully managed to ensure AI tools are validated using real world business data rather than based on experimental or hypothetical examples. The onus should be on the technology supplier, in partnership with the food business, to provide tools that are validated and fit for purpose. The FBO should be aware of the applications for which the tools have been developed and any limitations.

Case study discussions repeatedly highlighted that many AI systems available today are adapted from other domains and may not have been developed with food safety in mind. In Case Study 1 (risk assessment of manufactured foods), concerns were raised that generic AI systems might “lull users into a false sense of security” if validation was inadequate, while in Case Study 2 (third-party certification), participants emphasised that systems could misinterpret documentation unless they were trained on sector-specific, high-quality records.

These examples underline the need for transparent agreements between FBOs and suppliers that define how tools are validated, the data they are trained on, and the contexts in which they can or cannot be reliably used.

There is also a broader governance issue: food businesses remain legally accountable for food safety, but they may increasingly depend on AI suppliers for technical assurance. Case Study 3 (abattoirs) showed how AI could miss rare pathologies if suppliers failed to provide diverse training datasets, while Case Study 4 (ports) highlighted the importance of ongoing updates to keep pace with regulatory change. In both examples, weaknesses in supplier responsibility could directly undermine the ability of FBOs to demonstrate compliance. This risk underscores the importance of clear contractual frameworks that assign responsibility for validation, updates and transparency in system performance.

Going forward, closer collaboration between FBOs, AI providers, and regulators will be essential to avoid fragmented responsibility and ensure shared accountability. While ultimate legal responsibility for compliance cannot shift from the FBO, suppliers must be held to account for the quality, transparency, and robustness of their systems. Establishing common expectations for supplier validation, performance disclosure, and limitation reporting would not only protect businesses but also provide greater assurance to regulators and consumers. Without such safeguards, there is a danger that AI adoption could create new vulnerabilities in food safety rather than strengthening assurance.

- **Recommendation: Publish Guidance on Responsible Use of AI to Assure Food Safety and Regulatory Compliance**

Diversity and speed of introduction of AI tools and applications

AI is evolving rapidly, and its application is likely to change as capabilities mature, and costs fall. The extent and rate of this evolution is difficult to predict, making AI application highly dynamic. In addition, deployments in the food system are at an early stage, with many tools piloted in constrained settings rather than embedded into day-to-day operation and assurance. This makes real-world performance uncertain: behaviours observed in trials may not hold when systems face the variability of commercial operations, diverse datasets, and shifting standards. Against this backdrop, ongoing surveillance enables the FSA to observe how AI is actually being used, adapts over time and where new risks or opportunities emerge.   In addition, not all applications will be in the published scientific literature; awareness of the grey literature and business activities will also be essential. An additional target of a broader understanding of developments in AI use in food systems would be to ask if current supply chain standards are resilient to the possible use of AI to assist food fraud. The widespread availability of AI will undoubtedly attract criminals searching for ways to circumvent business food controls and regulatory checks. AI could assist in label counterfeiting, document fraud, and many help criminals find vulnerabilities in food systems.

Evidence from the case studies showed how unintended consequences may arise once systems are deployed. In Case Study 1 (risks for manufactured foods), participants cautioned that AI outputs can create a false sense of security if accepted uncritically, and that weaker operators might use AI to generate convincing risk assessments. In Case Study 2 (third-party certification), groups highlighted that multimodal document tools could take falsified or incomplete records at face value, producing authoritative-looking evidence lacking substance. Case Study 3 (abattoirs) emphasised the risk of performance drift and gaps around rare pathologies, arguing for long term trials at scale and continuous revalidation. Case Study 4 (ports of entry) raised concerns about large language models producing hallucinations and false positives, underscoring the need for monitoring and human challenge where results appear plausible but incorrect.

Because AI systems learn from and react to new data, one-off validation is not sufficient. Surveillance would allow the FSA to track adoption patterns (where, by whom, and for what decisions), watch key performance indicators over time (e.g. false-positive/negative rates, override and challenge frequencies, drift alerts), and identify signals of misuse or over-reliance (e.g. declining human verification, reliance on non-explained “black box” outputs without traceable evidence). It would also help the FSA spot data governance issues as they arise, such as inconsistent training data, poor provenance, or undocumented model updates, and target guidance or engagement accordingly.

- **Recommendation: Establish Ongoing Monitoring of AI Systems and Potential Impacts**

Data availability and quality as a major prerequisite for AI applications in food safety and authenticity assurance

High-quality, FAIR (Findable, Accessible, Interoperable, Reusable) data is fundamental to the development, exploitation and validation of AI systems in food safety, authenticity and assurance. The FSA should intensify its efforts to promote trusted food supply chain data sharing and alignment with recognised standards, ensuring datasets are protected from bias and drift, data provenance and ownership is clear, and regulatory consistency is maintained, including terminology, document formats and traceability frameworks. This builds on Science Council's WG4 report on data usage (Wolfe et al., 2020) and current FSA support for the Defra Food Data Transparency Partnership (FDTP) program. Alignment with UK food system vocabularies and record-keeping practices is essential to support transparent, auditable and fair AI behaviour. The barriers to AI captured in the Centre for Data Ethics and Innovation (2020) report including ethical and data barriers, are still highly pertinent. High quality data sharing and transparency is especially important as a means of detecting and preventing food fraud. The utility of AI as a tool to detect anomalies in supply chain data will be determined by the quality and accessibility of raw data from multiple sources. In summary, data access, quality and veracity is a pre-requisite for the successful application of AI tools by businesses and regulators to assure food safety and regulatory compliance. Increased data sharing would benefit all stakeholders and help prevent food fraud. Harmonized use of AI methodology across businesses and regulators would support consistent decision making and would build trust. There is a growing number of cases of cyber attacks on food businesses that have shut operations. In all cases, multiple sites and multiple enterprises were affected. In applying digital tools food businesses will need to look beyond their own organisation to guarantee cyber security.

The case study on border inspections (Case Study 4) demonstrated that without shared and standardised datasets, AI systems could not complete intended tasks (assuring food safety and provenance at borders). If AI systems are to deliver reliable outputs or even function, participants stressed that models must be trained on the full diversity of paperwork encountered at ports, including multilingual, handwritten and varying regulatory formats. Without this access, systems risk producing inaccurate or biased results, undermining both efficiency and trust. Missing data post training, when any AI is in operation, could render the system ineffective. Similar concerns were raised in Case Study 1 (risk assessment of manufactured foods), where participants warned that low-quality or incomplete input data could lead to poor outcomes, regardless of how sophisticated the AI appeared. Case Study 2 (third-party certification) also showed that AI tools may accept inconsistent records at face value unless data standards are robust. Taken together, these examples show that AI in food safety will not accomplish its intended tasks, or could deliver biased, incomplete, or misleading outputs, without access to high-quality, harmonised data. This makes data assurance and consistency not just a desirable feature but a fundamental prerequisite for meaningful AI deployment in food assurance.

At the same time, the FSA should recognise that not all FBOs, particularly SMEs, currently have equal access to the data infrastructure (internet / computer), expertise, or technical capacity required to fully benefit from AI. Generic AI tools may be accessible but of variable effectiveness for businesses due to these disparities. The FSA should develop an understanding of limitations and opportunities and ensure that efforts to improve data standards and digital capability are inclusive, enabling fair and proportionate adoption across the sector.

- **Recommendation: Promote Data Assurance, Validation, and Standards Alignment**

Food business vulnerability to AI products and applications that have not been rigorously tested or validated

The FSA should support the development of independent standards and validation mechanisms to ensure AI systems used in food safety are safe, reliable and fit for purpose. This may include the use of regulatory “sandboxes”, digital twins, or benchmarked synthetic data for independent testing prior to deployment. Such validation should assess key performance indicators, including false positives/negatives, hallucination risk, explainability, and consistency across environmental diverse conditions. An advantage of industry-driven standards and Codes of Practice is that best practice is likely to evolve at an accelerated pace requiring an agile process to capture new developments in a timely manner.

The case studies highlighted the need for systematic validation and shared standards and assurance frameworks to underpin trust in AI. In Case Study 2 (third-party certification), participants stressed that AI conclusions must be auditable and benchmarked against human evidence standards to be credible in regulated environments. Case Study 3 (abattoirs) reinforced this by emphasising the need for long-term efficacy trials and continuous revalidation to capture performance drift and rare pathologies. Case Study 3 also emphasised the need to quantify, as part of any standard, likelihood of false positive and negative results, these could have serious food safety consequences if not properly addressed. Whilst Case Study 4 (ports) showed that outputs must remain explainable and adaptable to changing regulatory standards if they are to be accepted by inspectors. Across these discussions, it was clear that without agreed standards and a common code of practice, AI systems risk uneven application, variable performance and erosion of trust. A coordinated framework would give businesses clarity on expectations and provide regulators with assurance that systems meet consistent, transparent benchmarks.

In parallel, the FSA should encourage the food industry, working with standards bodies, if necessary, and cross-government partners, to develop an industry-led Code of Practice for AI in food safety contexts. While not leading this directly, the FSA can act as a convenor and advisor to ensure alignment with regulatory expectations and consumer protection. The Code could draw on existing frameworks and help set clear expectations around data quality, governance, transparency and system robustness.

Coordination with wider government initiatives on AI assurance and standards will be important to ensure coherence across sectors while addressing the unique risks and regulatory needs of the food system.

- **Recommendation: Support the Development of Standards and an Industry-Led Code of Practice for Assuring AI in Food Safety**

Opportunities arising from similarities and synergies across different regulatory and policy domains

Given the wider implications of AI deployment across society, the FSA should engage with other relevant regulators, such as those within the Department of Science and Technology (DSiT), and appropriate international bodies to ensure coherence in governance and ethical standards. Lessons should be drawn from parallel domains (e.g. financial services, health diagnostics) where AI is being applied. Collaboration can also support benchmarking of assurance frameworks and AI auditability standards. The applications of AI in the domains of healthcare and clinical practice are developing at pace stimulating a number of commentaries and cautions on implications for regulatory compliance and quality assurance (e.g. Ong et al., 2025; Basubrin & Basubrin, 2025). It is imperative that regulators share knowledge of risks and opportunities.

The case studies made clear that AI in food safety is still in the early stages of adoption, with many tools at prototype or pilot level rather than scaled deployment. Case Study 4 (ports of entry) highlighted that without harmonisation of documentation standards across jurisdictions, AI could not deliver its intended function. Similarly, Case Study 3 (abattoirs) showed that validation of pathology detection tools requires not just technical testing but alignment with certification frameworks and inspector practice. These examples point to a broader context: many of the challenges facing the FSA mirror those in other domains, where regulators are facing similar questions of explainability, accountability and bias. It is therefore likely that the FSA can benefit from and contribute to this wider regulatory conversation rather than seeking to resolve these issues alone.

Engagement is also essential because AI applications are evolving rapidly, with new forms such as large language models and agentic AI emerging far faster than traditional regulatory processes can adapt. The case studies underscored the risk of unintended consequences: automation creep in abattoirs (Case Study 3), over-reliance on AI-generated certification packs (Case Study 2), or misplaced trust in unexplained outputs (Case Study 1). These risks highlight the importance of ensuring that the food system is aligned with cross-sector governance approaches that are developing in real time. A siloed approach could leave the FSA unprepared for the rapid diffusion of tools into food assurance that were originally designed for other industries.

Finally, engaging with broader regulatory and policy perspectives will help the FSA anticipate the legal and ethical shifts that are already beginning to shape AI deployment. The Law Commission (2025) has warned of liability gaps in autonomous and adaptive AI “where no natural or legal person is liable for the harms caused”, while international precedents such as the EU AI Act (2024) are setting new benchmarks for risk-based regulation. By working with other regulators and government departments, the FSA can ensure that food-specific concerns such as traceability, authenticity, and public health, are not overlooked in these wider frameworks. At the same time, this collaboration will give FBOs greater legal certainty and ensure that consumer trust in the UK food system is not undermined by inconsistent or fragmented approaches to AI governance.

- **Recommendation: Engage with Broader Regulatory and Policy Perspectives**

Human user understanding and application of AI outputs compared with those from conventional tools and advisors

To fully understand the potential risks and opportunities of AI deployment in the food system, the FSA should consider commissioning or supporting behavioural research, either directly by engaging the Advisory Committee for Social Science (ACSS) or with other funding agencies. While much attention has been given to the technical assurance of AI systems, less is understood about how human behaviours, such as cognitive bias, overreliance on automation, or misplaced trust in AI-generated outputs may impact food safety outcomes. While AI may lead to less demand for some human-actuated tasks, there may be an opportunity for assurance staff and regulators to focus on higher value activities such as targeted interventions during an inspection or audit. There may also be scope for skills development of the human workforce to enable earlier interventions to prevent food safety incidents leading to better consumer protection and lower business risks. However, vigilance may be needed to prevent essential skills deterioration particularly in small organisations.

The case studies revealed that the most significant risks may arise not from the technology itself, but from how people choose to work with it. In Case Study 1 (risk assessment of manufactured foods), participants noted that staff could become complacent, assuming AI-generated outputs were correct without carrying out independent checks, which risks embedding errors into safety plans. In Case Study 2 (third-party certification), auditors were concerned that over-reliance on AI-compiled assurance packs could discourage challenge. Case Study 3 (abattoirs) raised the possibility of inspectors gradually transferring too much responsibility to automated systems, leading to “automation creep” and even loss of critical skills over time. At ports of entry (Case Study 4), officials stressed that some users might avoid using complex AI tools altogether if they lacked training or trust, while others might even seek to game the system. These examples highlight the spectrum of human behaviours, from over-trust and passivity to avoidance and opportunistic misuse, that must be recognised to ensure AI supports, rather than undermines, food safety.

Research should explore how individuals across the food system (e.g. FBOs, inspectors, and consumers) engage with AI-generated information, including how trust is formed, when human oversight may weaken and how shortcuts may impact risk perceptions. Findings from sectors such as medicine and clinical decision support, where human-AI interaction is more advanced, could offer valuable insights.

Understanding these behavioural dynamics will help the FSA to develop more effective guidance, training, and governance approaches that account not just for the capabilities of AI systems, but also for the realities of human behaviour in operational settings.

- **Recommendation: Commission Research on Human Behaviour and Interaction with AI in Food Safety Contexts**

Conclusions

In this guide

In this guide

1. Food Standards Agency Science Council Report of Project ‘Artificial Intelligence Applications in Food Safety and Authenticity’
2. Authors, Acknowledgements and Declarations of Interest
3. Executive Summary
4. Recommendations to the FSA
5. Introduction
6. Methodology
7. Results and Discussion
8. Conclusions
9. Appendix A: Workshop Case Study Briefing Document
10. Appendix B: Workshop Participant List
11. Appendix C: Workshop Case Study Responses
12. References

Artificial intelligence is advancing at a pace that few other technologies have matched, with potential to reshape food safety and assurance into more predictive, transparent, and efficient systems. The case studies explored in this report illustrate both the promise and the complexity of this transformation. AI could allow hazards to be detected earlier, inspections to be scaled more effectively, and regulatory oversight to be enhanced by continuous monitoring of diverse data sources. It could enable more resilient, data-driven systems that support consumer protection while improving business efficiency.

At the same time, the case studies revealed that AI adoption in the food system is still at an early stage, with most applications tested only in pilots or controlled trials. This creates significant uncertainty about how tools will perform in real-world operations, where unintended consequences might emerge. Workshop participants raised concerns about issues such as performance drift, false assurance, over-reliance by users, and outputs that appear credible but lack traceable and explainable evidence. These concerns are not reasons to delay innovation, but they emphasise that vigilance, validation and human oversight are essential if AI is to strengthen, rather than weaken, assurance processes.

A consistent theme across the study was that AI should be treated as a decision-support tool, not a replacement for human accountability. Food business operators remain responsible for ensuring food safety and regulatory compliance and AI tools must be designed and governed in ways that make outputs explainable, auditable, and challengeable. Without these safeguards, the due diligence defences on which food law depends could be undermined. There remains significant uncertainty about how human behaviour will evolve in response to the growing use of AI in food safety and assurance contexts. New behavioural risks may emerge, including overreliance on AI outputs, reduced vigilance, or shifts in professional responsibility, all of which could introduce unintended impacts.

Another central finding is the importance of high-quality, harmonised data. AI cannot accomplish its tasks, or avoid bias, without access to diverse, representative, and trusted datasets. This is especially clear in contexts such as ports of entry and complex supply chains, where without shared and standardised records AI systems would be unable to function. Similarly, validation of AI in abattoirs depends on rich training datasets to capture rare or emergent pathologies. These examples highlight that data governance will be decisive in shaping the safe and fair use of AI.

Finally, the FSA cannot act in isolation. The rapid evolution of AI, including new forms such as large language models and emerging agentic AI systems, means that governance must be coordinated with wider policy and regulatory initiatives. Other sectors, from financial services to healthcare, are grappling with parallel questions of explainability, accountability and liability. Collaboration will ensure coherence across domains, avoid duplication of effort and help the FSA remain aligned with international best practice. The challenge ahead is not only technical, but social and ethical: embedding AI into food safety in a way that augments, rather than displaces, the human responsibility that underpins consumer protection and confidence in food law.

In closing, this report finds **no evidence at present that new food safety regulation is required** to address the use of AI. Existing frameworks should be sufficiently robust to encompass the application of currently known AI tools, provided human accountability remains central. However, this position is contingent on vigilance: AI adoption will continue to accelerate and evolve and, with it, new risks and opportunities will emerge. The FSA must therefore remain alert, adaptive, and proactive, maintaining ongoing surveillance, promoting trusted data standards, supporting validation and codes of practice, and engaging with wider regulatory debates.

AI has the potential to enhance the safety and regulatory compliance of the UK food system, but only if introduced carefully and responsibly. By adopting the recommendations set out here, the FSA can ensure that innovation is harnessed to protect consumers and strengthen public trust, while avoiding the risks of over-reliance, hype, or premature adoption.

Science Council Report of Project 'Artificial Intelligence Applications in Food Safety and Authenticity'

Appendix A: Workshop Case Study Briefing Document

In this guide

In this guide

1. Food Standards Agency Science Council Report of Project ‘Artificial Intelligence Applications in Food Safety and Authenticity’
2. Authors, Acknowledgements and Declarations of Interest
3. Executive Summary
4. Recommendations to the FSA
5. Introduction
6. Methodology
7. Results and Discussion
8. Conclusions
9. Appendix A: Workshop Case Study Briefing Document
10. Appendix B: Workshop Participant List
11. Appendix C: Workshop Case Study Responses

12. References

Food Standards Agency Workshop: Ensuring Safe and Trustworthy Application of AI in Food Safety and Assurance.

Preamble

As the use of artificial intelligence (AI) expands across the food sector, new opportunities and regulatory challenges are emerging. This workshop, hosted by the Food Standards Agency, brings together stakeholders from food safety, regulation, food production and supply and the AI community. Its objective is to explore how AI is, and might be, applied in the key domains of food safety and assurance. Using a series of case studies, participants will examine current and emerging applications ranging from risk assessment and certification to document inspection and to visual detection of defects. The workshop will consider the opportunities and challenges arising from AI technologies, where regulatory frameworks may need to evolve, what assurance mechanisms are required, and where potential gaps or risks may arise. The case studies will serve as a shared reference point for structured discussion, helping to identify areas where guidance, standards, or oversight could support safe and trustworthy deployment of AI across the food system.

Food Safety Management and Regulatory Compliance

Businesses employ a range of business processes and tools to ensure that food products placed on the market comply with relevant regulations and are safe to consume. Critical steps in such tools and processes require assurances that they work as designed and are fit for purpose. Such assurances will include, for example, implementation of recognized international, national and industry standards; reference to rigorous underpinning science; results of audits against such standards where applicable; evidence of operator competence where relevant; validation data (internal and external); supplier data (including traceability); customer and consumer complaints; contaminant and routine analytical and processing line data.

There is some degree of flexibility around how the above processes are implemented and ratified, as every business and product is different. However, any significant changes to the processes or procedures resulting from the introduction of AI technologies will require verification that the overall process of safety assurance and assessment of regulatory compliance is still operating as intended, and as specified in regulations across the food system. Such safeguards should encompass changes to physical steps such as harvesting and processing as well as data handling and analysis.

Case Study 1: AI Driven Safety and Regulatory Compliance Evaluation for Manufactured Foods

Food manufacturers developing complex, multi-ingredient products must conduct detailed safety evaluations and prepare food safety management plans to ensure products placed on the market are safe and comply with legislation. Key aspects of such evaluation and management planning include, for example, chemical and microbial safety, allergenicity, ingredient safety, and labelling accuracy and compliance. These assessments draw on an extensive and ever-evolving landscape of scientific, regulatory, and product data: from surveillance and monitoring data; validation data for processes and methods; shelf-life data; toxicological studies, historical data and published case studies; incidents data; and legal thresholds. The data are highly distributed, heterogeneous, and often unstructured. Complex supply chains and product recipes give rise to the risk of food fraud involving, for example, substitution with cheaper raw materials, or falsifying data regarding the origin or identity of the product. Effective supplier controls can lower the risk of food fraud. Databased approaches such as blockchains can protect the integrity of supplier data. In addition, numerous analytical tools are available. These include nucleic acid-based approaches such as DNA barcoding and chemical tools such as those based on spectroscopic fingerprints. Most such approaches generate large amounts of raw data that requires extensive analysis and expert interpretation before action can be taken. Some approaches are still experimental and therefore contentious.

To manage this complexity, manufacturers are likely to turn to **AI systems** to support early-stage safety, regulatory and labelling decisions. These systems may include:

- **Large Language Models (LLMs)** to process regulatory documents, scientific literature, and guidance.

- **Knowledge Graphs** and **ontologies** to map relationships between ingredients, allergens, and known risk pathways.

- **Multimodal AI** to integrate structured data (e.g., ingredient lists, concentrations, batch records) with unstructured text (e.g., literature or safety reports).

- **Rule-based** and **ML systems** to simulate or flag risks based on novel ingredient combinations or emerging science.

These tools can significantly accelerate product development while improving consistency and thoroughness. However, the use of AI for the assessment of regulatory compliance and safety evaluation raises critical assurance questions for the Food Standards Agency.

Key Questions for the workshop:

1. How can the FSA be assured that AI systems used for allergenicity and compositional risk assessments have accessed, interpreted, and applied the correct scientific and regulatory data across all relevant domains?

2. What standards should govern the transparency, traceability, and reproducibility of AI-derived risk assessments, particularly when used to justify labelling and safety decisions?

3. How do we validate that AI systems can identify emerging risks or uncommon ingredient interactions, and not just replicate existing knowledge—especially when considering public health risk?

Case Study 2: AI-Supported Data Pack Generation for Third-Party Certification and Assurance

Certification and assurance schemes require food producers and suppliers to demonstrate compliance with a wide range of standards across the entire food supply chain from farm to fork. Audits and inspections to ensure compliance with specified standards cover all aspects of food production, including food safety, traceability, production methods, worker safety, and environmental protection. Historically and currently, audits depend on the creation of detailed “data packs,” which draw on diverse, often fragmented, datasets across multiple systems and formats.

To automate the development of these “data-packs”, software developers are likely to explore the use of **multimodal AI systems.** These systems would integrate:

- **Natural Language Processing (NLP)** to interpret unstructured text such as process and quality records such as treatments on farm and on production line quality checks in manufacturing.

- **Optical Character Recognition (OCR)** to digitise and extract information from scanned or handwritten documents.

- **Tabular and Structured Data AI** to interpret spreadsheets, XML data, and inputs from farm and food manufacturing management software.

- **Document Question Answering (DocQA) and Retrieval-Augmented Generation (RAG)**, using large language models (LLMs), to automatically answer assurance protocol questions based on evidence extracted from multiple sources.

- **Rule-based Systems and ML Classifiers** to flag non-compliance, identify missing data, and suggest corrective actions.

The AI must be capable of reasoning across **heterogeneous, multimodal inputs**, often with incomplete, inconsistent, or domain-specific terminology. It must then map this information to the assurance scheme protocols—typically a complex, dynamic framework of hundreds of questions and compliance checks.

Key questions for the workshop:

1. How can we assure the robustness, consistency, and context-awareness of multimodal AI systems operating across diverse data sources?

2. What evidentiary standards must AI meet to ensure that its answers to assurance questions are auditable, transparent, and aligned with regulatory interpretation? How do these compare with current practices

for human inspectors?

3. How do we ensure that AI outputs can be validated, challenged, or corrected by human users—without undermining trust or introducing new risks?

4. How does such a system adapt to new regulations and standards over time, which are sometimes rapidly changing?

Case Study 3: AI-Assisted Detection of Infections and Other Pre/Post-Mortem Pathologies in UK Abattoirs

As part of an ongoing drive to enhance food safety and operational efficiency, software developers and equipment manufacturers are examining the use of machine learning (ML) to detect signs of infection, and even quality defects, through image recognition. Traditionally, this role is carried out by trained meat inspectors and official veterinarians, who visually assess carcasses for signs of disease or contamination.

AI systems developed by technology providers will deploy high-resolution imaging and ML models trained on thousands of annotated images to identify visual markers of infection and anomalies in real time. These systems would integrate:

- **Deep Learning** to detect visual anomalies or markers of pathology in real-time images or video streams.

- **Transfer Learning** to adapt pre-trained models for new or under-represented pathologies.

- **Edge AI** to enable on-device, real-time decision-making within abattoir environments.

In general, based on the known performance of "deep" learning detection of features in images, these systems are likely to have high potential and could well be scaled in industry at pace. In controlled environments and conditions, they might well demonstrate recall and precision rates comparable to those of human inspectors; where recall (true positives / (true positives + false negatives)) equates to a measure of detection rate and precision (true positives / (true positives + false positives)) to accuracy. However, variability in lighting, carcass presentation, environment, biological diversity, and rare pathogen manifestations are likely to remain key challenges. There are also important questions about the transparency and traceability of AI decisions, particularly in the context of regulatory compliance and public trust.

Key questions for the workshop:

1. How can regulators ensure that AI systems for infection detection achieve – and maintain – accuracy, recall, and stability at a level equivalent to or exceeding that of trained human inspectors?

2. What standards and validation processes should be established to evaluate the diversity and quality of training data, especially for rare or emergent pathogens where symptoms may not be well represented in existing datasets?

3. What criteria must be met before AI systems can be authorised for use in regulated environments, and how can ongoing performance be assured, particularly in terms of drift, bias, or unforeseen failures?

Case Study 4: AI-Powered Document Inspection at UK Ports of Entry

The UK Food Standards Agency along with software developers are exploring the use of AI systems, including large language models (LLMs), to enhance the inspection and verification of food import documentation at ports. These documents, ranging from health certificates and commercial invoices to packing lists and shipping manifests, are critical for ensuring food safety, regulatory compliance, and traceability of goods entering the UK. Traditionally, official controls involve officers manually reviewing these documents to assess conformity with safety standards and detect inconsistencies or fraudulent entries. This process can be time-consuming, especially under increased trade volumes and complex global supply chains.

An AI-based solution, incorporating document classification models, optical character recognition (OCR), and LLMs, could increase the productivity of frontline officers (e.g., freeing up time for more physical inspections and investigations). These systems can automatically extract key data points, cross-check documents for internal consistency, flag anomalies or incomplete submissions, and even interpret unstructured or multilingual content. LLMs, specifically, have shown promise in identifying subtle discrepancies in language, such as ambiguous product descriptions or suspicious edits.

To address these complexities, AI developers are likely to explore the use of multiple systems to automate document inspection. These systems would integrate:

- **Optical Character Recognition (OCR)** to digitise printed or handwritten documentation.
- **Document Classification Models** to identify and categorise incoming paperwork.
- **Large Language Models (LLMs)** for extracting, interpreting, and cross-validating information from unstructured or multilingual text.
- **Retrieval-Augmented Generation (RAG)** which only generates AI outputs that relate to retrieved external, or policy-specific, sources of truth, including published regulations and policy guidance.
- **Anomaly Detection Algorithms** to flag inconsistencies, duplications, or signs of fraud.

AI-enabled systems may well increase throughput with higher detection rates of noncompliant documentation, but questions remain as to how these systems can be assured, especially given the risk of hallucinations, bias, misinterpretation due to language subtleties, and over-reliance on AI outputs in critical decision-making.

Key questions for workshop:

1. How can we ensure the accuracy, reliability, and auditability of LLMs when used to assess official import documentation in regulated environments?

2. What safeguards are required to manage the risk of false positives or negatives, omissions, or AI-generated hallucinations, especially when

decisions impact food safety or border clearance?

3. How do we validate training data diversity and alignment with UK regulatory terminology, languages, and document formats to ensure equitable and robust performance?

4. What level of accuracy by an AI system is considered acceptable and would result in reducing burden on human inspectors? Should an AI system result be binary (pass or not) or should it provide a more nuanced output that includes reasons for an assessment?

AI Glossary

Anomaly Detection Algorithms

Techniques that identify outliers or irregularities in data; flagging errors, fraud, or noncompliance in complex documentation or operational workflows.

Deep Learning

An advanced form of machine learning using layered neural networks to recognise complex patterns; likely effective in tasks such as image recognition within carcass inspection.

Document Question Answering (DocQA)

An AI capability that allows systems to answer specific questions based on the contents of documents, useful for automating certification responses or audit checks.

Edge AI

AI that runs locally on devices rather than in the cloud, enabling real-time decision-making in operational settings like abattoirs or port inspections.

Food Assurance

Processes and schemes that provide verified confidence to consumers, regulators, and businesses that food has been produced, processed, and handled according to defined standards relating to safety, quality, animal welfare, and environmental impact.

Food Safety

Activities and measures aimed at protecting consumers from foodborne illnesses and contamination by ensuring that food is safe to eat. This includes the prevention, detection, and management of biological, chemical, and physical hazards throughout the food supply chain, in line with statutory requirements enforced by the Food Standards Agency.

Food Safety Management Plans

Structured documentation outlining processes, controls, and evaluations to ensure food products are safe for consumption and comply with regulatory standards. These plans incorporate assessments of chemical, microbial, and allergenic risks.

Knowledge Graphs

Structured networks that represent relationships between entities; such as ingredients, allergens, or contaminants, helping AI reason about risk pathways and regulatory linkages.

Large Language Models (LLMs)

AI systems trained on extensive text data that can interpret, summarise, and generate natural language. In food safety and assurance, they could be used to process regulatory documents, inspection records, and policy guidance.

Machine Learning (ML)

A core approach in AI where models learn from data to detect patterns and make predictions or decisions, used widely across safety assessment, document analysis, and image inspection.

Multimodal AI

AI systems capable of processing and integrating multiple data types, such as text, tables, images, and numerical values. This is especially useful in contexts where data is fragmented or presented in different formats.

Natural Language Processing (NLP)

A subfield of AI focused on understanding and interpreting human language potentially used to extract insights from farm records, safety reports, or multilingual documents.

Optical Character Recognition (OCR)

Technology that converts scanned, printed, or handwritten text into machine-readable data, allowing AI systems to work with legacy forms or paper-based documentation.

Retrieval-Augmented Generation (RAG)

A method where AI retrieves relevant external documents before generating a response, ensuring outputs are grounded in verifiable sources such as regulations or guidance notes. For example, if asked, “What are the UK import requirements for soft cheeses?”, a RAG-enabled system would first retrieve current FSA or border import guidelines, then generate a response based specifically on that content, thereby reducing the risk of error or hallucination.

Rule-Based Systems

AI tools that operate using fixed logic rules (e.g., “if X and Y occur, trigger a warning”), providing predictable outputs and supporting regulatory logic or protocol adherence.

Shelf-Life Data

Information generated from studies that determine how long a food product remains safe and

of acceptable quality under defined storage conditions. Critical for labelling, safety assessments, and regulatory compliance.

Third-Party Certification and Assurance

Independent verification processes where external bodies assess farms or food businesses against specific standards related to food safety, environmental protection, and good agricultural and processing practices.

Traceability

The ability to track the history, application, or location of a food product through all stages of production, processing, and distribution. Essential for rapid response to food safety incidents and regulatory compliance.

Appendix B: Workshop Participant List

In this guide

In this guide

1. Food Standards Agency Science Council Report of Project ‘Artificial Intelligence Applications in Food Safety and Authenticity’
2. Authors, Acknowledgements and Declarations of Interest
3. Executive Summary
4. Recommendations to the FSA
5. Introduction
6. Methodology
7. Results and Discussion
8. Conclusions
9. Appendix A: Workshop Case Study Briefing Document
10. Appendix B: Workshop Participant List
11. Appendix C: Workshop Case Study Responses
12. References

Appendix B: Workshop Participant List

Title	Full name	Affiliation	Position
Dr	Adam Cook	Food Standards Agency	Head of The Science and Research Unit
Prof	Ahmad Lotfi	Nottingham Trent University	Professor of Computational Intelligence
Mr	Alec Kyriakides	Independent	Consultant

Dr	Ayman Salama	Amazon Web Services	Senior Partner Solutions Architect
Ms	Cassandra Whelan	SUSTAIN AI CDT, University of Lincoln	Project Coordinator
Mrs	Catrin Hughes	MNGP Food Assurance	Technical Manager
Mr	Chris Gilbert-Wood	IFST, ex Two Sisters, Tech Director, ex M&S meat	Consultant
Ms	Claire Nicholson	Food Standards Agency Science Council	Member
Prof	Cronan McNamara	Crème Global & Adjunct Professor in University College Dublin	Founder and CEO
Ms	Dawn Welham	Aldi	QA Director
Prof	Emily Burton	Food Standards Agency Science Council	Member
Ms	Emma Hanby	Campden BRI	Associate Director – Research and Consulting
Prof	Georgios Leontidis	University of Aberdeen	Professor of Machine Learning
Mr	Imtiaz Shams	Flox	CEO

Mrs	Jacqueline Healing	Food Standards Agency Science Council	Member
Mr	James Whittaker	Boparan Food	Technical Director
Prof	Jeremy Graham Frey	School of Chemistry and Chemical Engineering, University of Southampton	Professor of Physical Chemistry
Dr	John Donaghy	Nestle	Global Head of Food Safety
Mr	John Figgins	LGC Group / BRCGS	Senior Technical Manager
Prof	John O’Brien	Food Standards Agency Science Council	Chair
Ms	Julia Hewerdine	Dawn Meats Group / Dunbia	Group Food Safety and Quality Projects Manager
Ms	Julie Pierce	Food Standards Agency	Director of Information and Science
Dr	Kathleen Mooney	Food Standards Agency Science Council	Secretary
Mr	Kiran Srimama	Google Cloud	Field Chief Technical Officer

Prof	Louise Manning	University of Lincoln	Professor of Sustainable Agri-food Systems
Prof	Mike Tildesley	Food Standards Agency Science Council	Member
Dr	Nick Birse	Queen’s University Belfast	PhD Supervisor
Mr	Paul Nunn	Food Standards Agency Science Council	Secretariat Lead
Prof	Peter Gregory	Food Standards Agency Science Council	Member
Dr	Phil Hadley	International Meat Secretariat (IMS)	Secretary General
Prof	Po Yang	The University of Sheffield / School of Computer Science	Professor of Pervasive Intelligence
Prof	Rich Smith	Food Standards Agency Science Council	Member
Prof	Rick Mumford	Food Standards Agency	Deputy Chief Scientific Advisor
Mr	Robert Nugent	Prosur (Productos Sur)	Technical sales manager
Prof	Roger Maull	University of Exeter	Professor of Management Systems
Ms	Sara Mortimore	Ex-Walmart	Consultant

Prof	Simon Parsons	University of Lincoln	Professor of Machine Learning
Prof	Simon Pearson	Food Standards Agency Science Council	Member
Mr	Simon Thorpe	Red Tractor (Assured Food Standards)	Senior Technical Manager
Prof	Steven Cummins	Food Standards Agency Science Council	Member
Miss	Victoria Balch	Food Standards Agency	Secretariat
Mx	Villanelle O’Reilly	UKRI SUSTAIN AI CDT (Hosted at the University of Lincoln, School of Engineering and Physical Sciences)	PhD Student

Appendix C: Workshop Case Study Responses

In this guide

In this guide

1. Food Standards Agency Science Council Report of Project ‘Artificial Intelligence Applications in Food Safety and Authenticity’
2. Authors, Acknowledgements and Declarations of Interest
3. Executive Summary
4. Recommendations to the FSA
5. Introduction

6. Methodology
7. Results and Discussion
8. Conclusions
9. Appendix A: Workshop Case Study Briefing Document
10. Appendix B: Workshop Participant List
11. Appendix C: Workshop Case Study Responses
12. References

Case Study 1: AI-Driven Safety and Regulatory Compliance Evaluation for Manufactured Foods

Question 1. How can the FSA be assured that AI systems used for allergenicity and compositional risk assessments have accessed, interpreted, and applied the correct scientific and regulatory data across all relevant domains?

“AI is a tool for helping-you need to know what questions to ask it.”

“There needs to be some verification that what AI is providing is correct.”

“AI gives you an answer and lulls into false sense of security. People do not give enough thought to the answer.”

AI should be appraised in the same way as any other tool used in food safety assurance. The same safeguards and assurance processes should apply as with non-AI systems. The overall process would not fundamentally change even if AI were deployed widely. Measurement of effectiveness would be through appropriate output and outcome measures. Any negative changes picked up during inspections should be raised with the FBO who should be in a position to provide a satisfactory explanation. If not, it could be a red flag for further investigation.

While processes may not need to change significantly, data quality, accessibility, and sharing must be addressed. Also, AI could support oversight and insights not possible within the current process so processes and procedures might change incrementally in response to opportunities. Opportunities may present that allow for better separation of purely food safety issues versus broader food quality parameters that are not of primary concern to FSA. It was emphasized in the Plenary Session that current UK food regulations are robust and do not require major revision. A complementary perspective was the suggestion that it would be a missed opportunity to limit AI to performing current tasks.

Automation of routine checks is potentially a major opportunity. This might free human agents to devote time to more valuable tasks such as investigating defects or determining root causes of failures.

In well-run food businesses, the addition of AI tools can support and potentially improve food safety controls and assurance. There is a danger in some of the less well-run operations that adoption of AI solutions has the potential to hide poor management by, for example, generating documents and plans that look very good but which are not fully implemented or understood in the organization.

Question 2. What standards should govern the transparency, traceability, and reproducibility of AI-derived risk assessments, particularly when used to justify labelling and safety decisions?

"If had recall – the authority comes in and asks what detection systems do you have in place. I explain to the authorities I use an AI trained x-ray system. The FSA will want to know how the system has been trained."

"Transparency and traceability is OK. Reproducibility... this is the AI business, not the FSA's responsibility."

There was a lack of consensus as to whether new standards are necessary governing, inter alia, transparency, traceability and reproducibility of AI derived risk assessments. If standards were deemed to be necessary, consideration would need to be given to timing, as tools and applications are evolving. The introduction of unnecessary additional standards may well hamper innovation. While standards might be a consideration for future development, current use of AI in risk assessments should be date-stamped as AI will evolve over time.

Participants consistently stressed the importance of maintaining a 'human-in-the-loop' approach. Similarly, AI was seen as a supporting tool (e.g. a co-worker) but not replacing human oversight in critical decisions.

Data are key. More data sharing will be needed. Implementation of AI tools would be supported by high quality information in digital form. Digital transformation of processes is needed. Both of these points have broad implications: data sharing requires trust among different stakeholders in the food system and, at least in some instances, there is a lack of trust. Digital data may require investment and adoption of new technologies. Both changes need concerted action; there are examples where opportunities for digital data sharing were undermined by stakeholders still using legacy systems.

Data governance is needed; data integrity was raised several times. Low input data quality has the potential to lead to poor results and decisions and FBOs should be aware of this.

FSA could support by providing guidance to businesses on the deployment of AI tools. Participants highlighted the opportunity for training; however, training of FBOs would not be for FSA to do.

The term “compliance skill” was used to describe the need of FBOs to understand the food system. In evaluating the significance of any process changes, an understanding of the basis of current controls is important. Microbusinesses and SMEs were highlighted as needing tailored support due to limited resources and technical expertise.

There may be a need to provide training to food officials to ensure awareness of the tools and to ensure they understand the implications. Introduction of AI tools may generate more complexity from a governance perspective.

Question 3. How do we validate that AI systems can identify emerging risks or uncommon ingredient interactions, and not just replicate existing knowledge—especially when considering public health risk?

“AI can be used to test the risk and the impact of the risk and then model it.”

Emerging risks identified by AI should be validated using methods consistent with current practice. However, AI could be used to rapidly simulate different scenarios in a way that would be difficult or impossible currently.

Case Study 2: AI-Supported Data Pack Generation for Third-Party Certification and Assurance

This section presents an in-depth analysis of the challenges and opportunities in applying AI to automate and augment data pack generation for third-party food certification and assurance. These packs often consist of mixed-format documentation, scanned reports, sensor outputs, handwritten logs; all of which must be synthesized to inform certification decisions. AI systems can offer real-time assistance, retrospective pattern analysis, and expanded coverage for audits, but only if they are implemented within robust governance structures. Manning et al. (2022) argue that responsible AI must be traceable, explainable, accountable, and accessible. These principles must not only be present at deployment but also evolve with continued operation. AI systems deployed in assurance contexts must remain open to human challenge and reinterpretation throughout their lifecycle.

1. How can we assure the robustness, consistency, and context-awareness of multimodal AI systems operating across diverse data sources?

"AI may take falsified records at face value. No change in pen used between shifts could be a red flag.";

"AI can look across the supply chain... auditors only able to look at a small selection of data."

Participants highlighted the complexity of interpreting mixed-format documentation, especially when human behaviours, workarounds, or anomalies exist. AI tools must go beyond surface extraction to understand context. These tools must not only process multimodal inputs, but also infer operational context; when and where a data recording was written, how it aligns with sensor data, and who entered it. Workshop participants suggested that AI models should be trained on datasets reflecting the full range of operational variability, including uncommon edge cases. The goal is not just information extraction but evidence synthesis; an ability to build narratives from fragmented sources, really as a human auditor would.

"Systems that only read structured fields miss the nuance of how and why data was recorded. AI should learn to reason, not just match templates."

Participants recommended embedding ‘context windows’ into model architectures, enabling models to interpret temporal and spatial relationships within data sets. By assessing the sequence of inputs and recognizing cross-modal inconsistencies, AI can begin to emulate the human capacity for contextual inference. This would help identify when data appears artificially consistent or repetitive, common potential evidence for fraud, and better support assurance judgments that rely on human intuition.

2. What evidentiary standards must AI meet to ensure that its answers to assurance questions are auditable, transparent, and aligned with regulatory interpretation? How do these compare with current practices for human inspectors?

"Two auditors may have a different conclusion both looking at the same factory."

"Two auditors may reach different conclusions; a model must be transparent enough to support its claims with evidence."

The workshop reflected uncertainty about the evidence base required to support AI-generated conclusions. AI must codify and consistently apply regulatory standards. Where interpretation varies by scenario, AI should defer to humans or be configured with confidence thresholds. Explainability must be built into assurance tooling, allowing users to interrogate not only conclusions but also their evidentiary base. Proposed measures included visual traceability maps, structured logic chains, and metadata summaries for each judgement.

"If a human auditor made a claim, they'd have to back it up with evidence. We should expect the same of an AI model."

Participants called for an agreed validation standard that would guide how AI models present audit evidence. This might involve linking each judgement to its data source and identifying assumptions the model made during inference. By mirroring human evidence standards, AI systems could provide equivalent or superior levels of audit traceability. Some advocated for regulatory 'assurance blueprints' that define which decisions must remain under human control.

3. How do we ensure that AI outputs can be validated, challenged, or corrected by human users—without undermining trust or introducing new risks?

"A transparent system builds confidence, built by competent people, not just industry or regulators."

Participants emphasized the importance of embedded challenge mechanisms and proportional validation frameworks. AI interfaces should allow users to flag unexpected outputs, request supporting evidence, and provide corrective feedback. This feedback must be traceable, auditable, and, where appropriate, incorporated into model refinement. Suggestions included escalation paths, feedback loops for retraining, and role-based permissions for challenging AI conclusions.

"It's not just having a 'challenge button', the system should also log how often and by whom it's used."

A system's ability to maintain trust depends on its capacity to support structured disagreement. Attendees proposed audit logs of all user interactions, including overrides and challenges, to track patterns and flag areas of recurring concern. Such logs could inform policy refinement and potentially identify systematic weaknesses in model performance or user interpretation.

4. How does such a system adapt to new regulations and standards over time, which are sometimes rapidly changing?

"If you want AI answers to be the truth, you need to rebuild and retrain models after every regulatory{standards} update."

Participants underscored the need for model agility and standards integration. Proposals included the release of machine-readable updates by standards bodies, co-designing rules with developers, and maintaining version-controlled models linked to evolving standards. Standards sandboxes were suggested for testing tools under future state conditions without disrupting assurance workflows.

"The challenge isn’t just that standards change, it’s that interpretations change too."

Attendees proposed collaborative working groups to interpret changes in practice and provide pre-configured rule updates to developers. Others supported embedding legal logic frameworks in assurance platforms that alert users when changes in standards affect data interpretation or system confidence. In fast-moving standards environments, continuous professional development will be essential, not only for AI systems, but for human overseers as well.

General Reflections and Conclusion

The use of AI in third-party certification presents substantial opportunities for scaling audit capacity and enhancing traceability. However, systems must be grounded in principles of transparency, accountability, and human oversight. Throughout the workshop, participants consistently emphasized the importance of hybrid decision-making models where AI supports rather than replaces human inspectors.

A key takeaway was the importance of building trust not only in the system's technical performance but also in its governance and resilience. Participants warned against over-reliance on AI in contexts where the human touch/ auditor skill remains critical.

Going forward, effective deployment will require close collaboration between system designers, certifiers, standards bodies and end users. Robust training, version control, explainable outputs, and consistent human-in-the-loop validation processes will be critical to ensuring AI supports food assurance safely and effectively.

Participants also emphasised the scale advantage AI offers. Unlike traditional auditors who sample selectively, AI systems can scan entire datasets across supply chains in real-time, identifying continuous trends and anomalies with greater precision. This capacity can support auditors by flagging areas requiring human interpretation. However, the utility of such wide-ranging surveillance is conditional upon maintaining data quality and interpretability across heterogeneous systems.

There was concern that human audit consistency varies significantly. AI could bring a more standardised evaluation baseline, but only if benchmarked effectively against multiple auditors and real-case discrepancies. Transparency in how AI reaches conclusions was considered critical, particularly where automated systems supplement or replace human decision-making.

Several contributors underscored that human audits are not infallible, and the benchmark for AI should be parity with, or improvement on, current outcomes. The emphasis was not on perfection, but on traceability and the availability of mechanisms to flag, interrogate, and override AI errors. Ensuring that the supporting evidence remains auditable was seen as key to maintaining public and regulatory trust.

Experts warned that systems adapting to shifting standards and regulations must undergo rigorous revalidation. Quick updates, while technically feasible, may lead to untrustworthy outputs if not properly verified. Establishing protocols for AI lifecycle management, including regular re-training and validation cycles, was considered essential to sustaining compliance over time.

Case Study 3: AI-Enabled Pathology Detection in Abattoirs

This section presents an analysis of the challenges, opportunities, and regulatory implications of integrating AI systems into meat inspection processes at UK abattoirs. Based on findings from the FSA workshop and grounded in responsible AI principles (Manning et al., 2022), the discussion focuses on the use of AI for real-time detection of pre- and post-mortem pathologies. The section addresses critical regulatory questions relating to accuracy, data quality, validation, and governance frameworks, while emphasising the importance of human oversight in high-risk environments. Manning et al. (2022) highlight that responsible AI must be designed to be traceable, explainable, accountable, and accessible, with these properties evolving over time as systems are deployed and interact with complex environments. They argue that AI must remain open to human challenge and reinterpretation, not only at design but throughout its lifecycle, reinforcing the need for transparency and continuous learning.

How can regulators ensure that AI systems for pathology achieve, and maintain, accuracy, recall, and stability at a level equivalent to or exceeding that of trained human inspectors?

"Systems need to be tested over a long enough period to cover rare pathologies and ensure performance in operational settings." Ensuring AI systems match or surpass human performance requires validation under real-world conditions. During the workshop, participants emphasised the need for longitudinal trials, ideally spanning six months or more, during which AI decisions are shadowed and reviewed by human inspectors. Such trials must encompass a diverse range of carcasses and environmental conditions to simulate operational variability.

"Performance should be monitored not just at launch but continually, systems need metrics tied to human oversight." AI model performance should be benchmarked against human inspection rates for both true positives and false negatives. Importantly, the system should provide confidence scores and justifications for its outputs, thereby supporting transparency and enabling human reviewers to make informed decisions. Participants supported a ‘closed-loop’ assurance cycle, in which real-world outcomes inform model retraining. In terms of standards, routine recalibration and scenario testing would be mandated as part of system governance, especially when model performance falls outside acceptable thresholds. Continuous monitoring, supported by a secure audit trail would ensures that AI does not degrade over time or drift away from its validated state.

What standards and validation processes should be established to evaluate the diversity and quality of training data, especially for rare or emergent pathogens where symptoms may not be well represented in existing datasets?

“There is a need for richer training data, particularly around rare or emergent conditions, with variation in lighting, angle, and pathology type”. Training data quality was highlighted as a critical foundation for AI success. Diverse, high-resolution imagery from multiple processing sites is required to reflect the full range of carcass conditions, including lighting variability, camera angles, and rare pathology types. Participants emphasised that without rich training datasets, AI tools could fail to detect uncommon or emerging conditions, thereby compromising food safety. Workshop recommendations included establishing a centralised and anonymised dataset repository. This would allow for consistent evaluation of AI tools and potentially facilitate data sharing by technology developers.

"Models should be tested against diverse and anomalous datasets to validate robustness in the real world." Annotated datasets should include metadata on inspection context, species, and confirmed pathology outcomes to support meaningful benchmarking. Validation should include stress-testing models with deliberately ambiguous or noisy inputs to assess robustness. Standards should mandate the inclusion of under-represented conditions and specify acceptable minimum data volumes for new model releases. The aim is to ensure representativeness, fairness, and performance consistency across use contexts. Models should be tested against diverse (including anomalous) datasets to validate robustness in the real world.

What criteria must be met before AI systems can be authorised for use in regulated environments, and how can ongoing performance be assured, particularly in terms of drift, bias, or unforeseen failure?

"We need to monitor for model drift, bias and maintain performance, unexpected failures need detection mechanisms." Authorisation of AI systems in regulated environments requires the development of a formal certification framework. Such a standards framework should define pre-market testing conditions, performance thresholds, data traceability requirements, and human-AI interface standards. Workshop participants proposed a tiered deployment model beginning with supervised trials, followed by staged rollout with increasing autonomy. Each deployment stage would require predefined success metrics, potentially with regulatory oversight at every phase.

"There should be independent checks for fairness, especially across sites using different camera technologies or lighting setups." To guard against bias and drift, systems must incorporate self-monitoring tools and notify operators if key indicators deviate from expected norms. Manning et al. (2022) suggest lifecycle governance approaches, including automated alerts and scheduled revalidation. Independent auditors could be employed to conduct performance and bias assessments at regular intervals. Any material changes to AI architecture or training data should trigger re-validation/certification. There should be independent checks for fairness, especially across sites using different animals, batches, camera technologies or lighting setups.

General Reflections and Conclusion

"We must define where human judgement must override AI, particularly for ambiguous or edge cases." Beyond technical performance, the ethical dimensions of AI use in abattoirs were a recurring concern. Participants stressed that AI should not become a substitute for human judgement in food safety. The ‘two in a box’ model, wherein AI works in tandem with a human inspector, emerged as a governance mechanism. This model balances efficiency with oversight, ensuring that AI augments rather than replaces human expertise. Participants also warned of ‘automation creep’, where AI systems gradually assume decision-making roles without explicit governance sanction. To mitigate this, policy should consider clearly defined thresholds beyond which only a human can make or validate decisions. Furthermore, retraining and upskilling programmes are vital to ensure inspectors remain capable of effective oversight in increasingly digital inspection environments.

The integration of AI into abattoir inspections promises substantial gains in safety, efficiency, and consistency. However, realising these benefits safely depends on transparent governance, robust validation, ethical safeguards, and a commitment to sustaining human expertise. AI must serve as a co-pilot, not a replacement, within the well-defined and understood regulatory frameworks. We must define where human judgement must override AI, particularly for ambiguous or edge cases.

Case Study 4 – AI-Powered Document Inspection at UK Ports of Entry

Background

The UK Food Standards Agency along with software developers are exploring the use of AI systems, including large language models (LLMs), to enhance the inspection and verification of food import documentation at ports. These documents, ranging from health certificates and commercial invoices to packing lists and shipping manifests, are critical for ensuring food safety, regulatory compliance, and traceability of goods entering the UK. Traditionally, official controls involve officers manually reviewing these documents to assess conformity with safety standards and detect inconsistencies or fraudulent entries. This process can be time-consuming, especially under increased trade volumes and complex global supply chains.

An AI-based solution, incorporating document classification models, optical character recognition (OCR), and LLMs, could increase the productivity of frontline officers (e.g., freeing up time for more physical inspections and investigations). These systems can automatically extract key data points, cross-check documents for internal consistency, flag anomalies or incomplete submissions, and even interpret unstructured or multilingual content. LLMs, specifically, have shown promise in identifying subtle discrepancies in language, such as ambiguous product descriptions or suspicious edits.

1. How can the accuracy, reliability, and auditability of LLMs be assured when used to assess official import documentation in regulated environments?

"LLMs are helpful but only partly, looking to narrow it down for greater degree of success."
"Need to monitor ongoing performance of the system. People are always looking for new ways to bypass systems."
"Using humans in the loop to improve."
"AI can look for spikes and anomalies that flag the need for further investigation."

LLMs are a new 'cultural technology' that allows individuals to take advantage of collective knowledge, skills and information accumulated through human history. Implicit elements of intelligence that are largely omitted from current models include: wisdom and judgement developed through experience; creative insight that transcends pattern recombination; intuitive understanding that cannot be verbalized; embodied knowledge learned through physical interaction; and self-awareness and metacognition. Many of these omissions featured in the workshop discussion.

The current priority at ports is food safety but if OCR and LLMs can remove some of the burden of risk assessment from inspectors then more time could be spent on ensuring authenticity, and preventing fraud and smuggling. The purpose/question underpinning document inspection needs to be clear and judgements about acceptability of the paperwork need to be consistent. The system needs to cope with documents from a variety of sources including printed, handwritten and scanned material of low resolution. Assessments vary in complexity. In addition to registering right/wrong responses to questions, the system could also look for anomalies (e.g. a spike of a product entering from a region/country not seen before) requiring further investigation. The ongoing performance of the system needs to be monitored and improved in the light of experience.

2. What safeguards are required to manage the risk of false positives or negatives, omissions, or AI-generated hallucinations, especially when decisions impact food safety or border clearance?

"It's not just having a 'challenge button', the system should also log how often and by whom it's used."
"We still need humans to sample even when it looks like it's working."

"Confidence levels should be known"
"Explainability must be built into assurance"

Current AI systems are designed to produce an answer with hallucination occurring when the system generates plausible but incorrect information. Continuous checking, retraining and validation are necessary to mitigate these risks.

A potential benefit of AI systems is that an audit approach could be employed giving a feel for intent and corporate activity of the organization. While some decisions are binary, others are more nuanced with human judgement required. Confidence levels and contextual analysis could be part of the AI output to inform human decisions.

The system should be gradually introduced with clear guidance on the roles of machines and humans in decision making. A continuous learning approach should be adopted with the role of humans changing as AI systems learn. AI systems can be used to detect anomalies leading to human interventions to mitigate errors in decision-making.

3. How to validate training data diversity and alignment with UK regulatory terminology, languages, and document formats to ensure equitable and robust performance?

"You never get perfect data. You get imperfect data and work with what you have."
"These systems must align with our forms, terms and regulation and be updated when regulations change."
"If you want AI answers to be the truth, you need to rebuild and retrain models after every standards update."

Participants stressed that AI models used at UK ports must be trained on datasets that reflect the full diversity of documentation encountered, including non-English content, handwritten submissions, and forms generated under differing regulatory regimes. Without this diversity, models may perform well in ideal cases but fail in realistic or marginal ones.

Alignment with UK-specific terminology and standards was seen as essential. This includes updating models when guidance changes, and fine-tuning LLMs on validated local data. Validation should therefore be continuous. In practice, this may include shadow testing (comparing AI outputs against human decisions), spot audits, and real-time confidence scoring. Ultimately, data diversity and regulatory alignment should be treated as core governance requirements rather than optional refinements.

4. What level of accuracy by an AI system is considered acceptable and would result in reducing burden on human inspectors? Should an AI system result be binary (pass or not) or should it provide a more nuanced output that includes reasons for an assessment?

“AI should be nuanced and leave the decision up to the human.”
“You need a reason why something isn’t allowed in.”

Participants agreed that no fixed accuracy threshold would universally justify removing human oversight. Rather, acceptable performance must be judged in context, by comparing AI output to current inspection accuracy and by determining how the tool complements rather than replaces human judgement.

Binary outputs may be useful for some high-confidence cases but were generally seen as insufficient for complex scenarios. A tiered system was suggested, where AI models provide a confidence score and rationale, enabling human inspectors to decide whether further investigation is needed. Such systems would allow inspectors to prioritize workloads more effectively, directing attention to ambiguous or borderline cases.

Ultimately, AI should enable better decisions, not faster errors. Participants emphasized that clarity of explanation, traceability of decision logic, and ongoing human review would be key determinants of a model’s fitness for purpose.

References

In this guide

In this guide

1. Food Standards Agency Science Council Report of Project ‘Artificial Intelligence Applications in Food Safety and Authenticity’
2. Authors, Acknowledgements and Declarations of Interest
3. Executive Summary
4. Recommendations to the FSA
5. Introduction
6. Methodology
7. Results and Discussion
8. Conclusions
9. Appendix A: Workshop Case Study Briefing Document
10. Appendix B: Workshop Participant List
11. Appendix C: Workshop Case Study Responses
12. References

Basubrin, O. and Basubrin, O., 2025. Current status and future of artificial intelligence in medicine. Cureus, 17(1). e77561.doi:10.7759/cureus.77561 https://doi.org/10.7759/cureus.77561

Benefo, E.O., Karanth, S. and Pradhan, A.K., 2022. Applications of advanced data analytic techniques in food safety and risk assessment. Current Opinion in Food Science, 48, p.100937.
https://doi.org/10.1016/j.cofs.2022.100937

Centre for Data Ethics and Innovation (2020), Independent report.

AI Barometer 2020, https://www.gov.uk/government/publications/cdei-ai-barometer

European Parliament and Council (2024) Regulation

(EU) 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), Official Journal of the European Union, OJ L 2024/1689, 12 July. Available at: http://data.europa.eu/eli/reg/2024/1689/oj.

Huckins, G. (2025), Are we ready to hand AI agents the keys. MIT Technology Review, 2025.

Kabir, M.A., Lee, I., Singh, C.B., Mishra, G., Panda, B.K. and Lee, S.H., 2025. Detection of Mycotoxins in Cereal Grains and Nuts Using Machine Learning Integrated Hyperspectral Imaging: A Review. Toxins, 17(5), p.219. https://doi.org/10.3390/toxins17050219

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems (pp. 1097-1105).

Law Commission (2025) AI and the Law: Discussion Paper, Law Commission of England and Wales, 21 July. Available at: https://cloud-platform-e218f50a4812967ba1215eaecede923f.s3.amazonaws.com/uploads/sites/54/2025/07/AI-paper-PDF.pdf (Accessed: 7 September 2025).

Li, Z., Chapin, A., Xiang, E., Yang, R., Machado, B., Lei, N., Dellandrea, E., Huang, D. and Chen, L., 2025. Robotic Manipulation via Imitation Learning: Taxonomy, Evolution, Benchmark, and Challenges. arXiv preprint arXiv:2508.17449.

Manning, L., Brewer, S., Craigon, P.J., Frey, J., Gutierrez, A., Jacobs, N., Kanza, S., Munday, S., Sacks, J. and Pearson, S. (2022). Artificial intelligence and ethics within the food sector: Developing a common language for technology adoption across the supply chain. Trends in Food Science & Technology, 125, pp.33-42. https://doi.org/10.1016/j.tifs.2022.04.025.
https://doi.org/10.1016/j.tifs.2022.04.025

Munck, N., Leekitcharoenphon, P., Litrup, E., Kaas, R., Meinen, A., Guillier, L., Tang, Y., Malorny, B., Palma, F., Borowiak, M. and Gourmelon, M., 2020. Four European Salmonella Typhimurium datasets collected to develop WGS-based source attribution methods. Scientific data, 7(1), p.75. https://doi.org/10.1038/s41597-020-0417-7

Ong, J.C.L., Ning, Y., Liu, M., Ma, Y., Liang, Z., Singh, K., Chang, R.T., Vogel, S., Lim, J.C., Tan, I.S.K. and Freyer, O., 2025. Regulatory science innovation for generative AI and large language models in health and medicine: a global call for action. arXiv preprint arXiv:2502.07794.

OpenAI (2022) 'Introducing ChatGPT', OpenAI, 30 November. Available at: https://openai.com/index/chatgpt/(Accessed: 24 August 2025).

Pennells, J., Watkins, P., Bowler, A.L., Watson, N.J. and Knoerzer, K., 2025. Mapping the AI Landscape in Food Science and Engineering: A Bibliometric Analysis Enhanced with Interactive Digital Tools and Company Case Studies. Food Engineering Reviews, pp.1-25.
https://doi.org/10.1007/s12393-025-09413-w

Qian, C., Fan, X., Shang, Y., Xu, X. and Wang, S. (2023) 'How can AI help improve food safety?', Annual Review of Food Science and Technology, 14(1), pp. 517-538. https://doi.org/10.1146/annurev-food-060721-013815
https://doi.org/10.1146/annurev-food-060721-013815

Shoham, Y. (2025), Don't let hype about AI agents get ahead of reality. MIT Technology Review, July 3 2025.

Tao, D., Zhang, D., Hu, R., Rundensteiner, E. and Feng, H., 2021. Crowdsourcing and machine learning approaches for extracting entities indicating potential foodborne outbreaks from social media. Scientific reports, 11(1), p.21678.
https://doi.org/10.1038/s41598-021-00766-w

Wolfe, P. (2020). Science Council Working Group on Data Usage and Digital Technology. (2020). Final Report to the Food Standards Agency (FSA). Working Group 4, June 2020,
https://www.food.gov.uk/sites/default/files/media/document/fsa-20-08-07-annex-report-from-the-science-council-working-group-on-data-usage-and-digital-technology.pdf. Accessed 1 Sept 2025

Zhao, Z., Wang, R., Liu, M., Bai, L. and Sun, Y., 2025. Application of machine vision in food computing: A review. Food Chemistry, 463, p.141238.
https://doi.org/10.1016/j.foodchem.2024.141238