Prison Reform Trust response to the Joint Committee on Human Rights' inquiry on Human Rights and the Regulation of AI – September 2025

The Prison Reform Trust (PRT) is an independent UK charity working to create a just, humane and effective penal system. We do this by inquiring into the workings of the system; informing prisoners, staff and the wider public; and by influencing Parliament, government and officials towards reform. The Prison Reform Trust provides the secretariat to the All Party Parliamentary Penal Affairs Group and has an advice and information service for people in prison.

The Prison Reform Trust's main objectives are:

- reducing unnecessary imprisonment and promoting community solutions to crime
- improving treatment and conditions for prisoners and their families
- promote equality and human rights in the criminal justice system.

www.prisonreformtrust.org.uk

Introduction

1. Drawing on our expertise in prisons, we focus our response on the development and adoption of AI and its potential human rights implications for people in the criminal justice system—particularly those in custody—and on the call for evidence questions that best address these issues.

2. As the Ministry of Justice is one of a small number of government departments with published plans for the development and adoption of AI, and the potential implications for ECHR Articles 2 (right to life), 3 (freedom from torture and inhuman or degrading treatment), 5 (right to liberty and security), and 14 (protection from discrimination) we recommend that the committee requests oral evidence from ministers and/or officials within the department to inform its inquiry.

3. It's to its credit that the Ministry of Justice is attempting to set out how it will approach the design, development and adoption of AI technologies to address the considerable challenges the department faces. The Prison Reform Trust is not anti-technology or innovation, but it is important to fully consider the implications and potential risks, when the increased sophistication of generative AI models and development timescales are rapidly accelerating. We welcome the consideration that the committee is giving to this matter.

4. Previous examples have shown that the development and adoption of tools with little communication or publicly available information can sow mistrust and

misunderstanding.[1] It is important therefore that a culture of justified trust[2] is developed between HM Prison and Probation Service; people in prison and on probation; the organisations who support them; and other stakeholders.

5. It is right that the Ministry of Justice is looking to provide technological solutions to meet the challenges it currently faces to reduce record-high levels of violence, death and self-harm in its prisons. The selective and carefully managed deployment of AI tools could provide frontline staff with the time they need to focus on rebuilding positive relationships with prisoners, and alleviate the issues which have undermined prison safety over the last decade and a half. The restoration of effective relationships and trust between prisoners and staff could also improve the fidelity of information and intelligence within prisons.

6. But there remains a significant distance to travel to reach this point, and there is already well-established evidence on effective ways to improve safety which don't require the development of new algorithms to attempt to predict risk.[3]

The use of algorithmic tools and AI in the prison system

7. Algorithmic tools are already used by professionals within the criminal justice system in England and Wales to support decision making at almost every stage. These range from patrol planning, crime discovery, and risk prediction by police forces; to informing court sentences; and assessing prisoner security categorisation[4], likelihood of re-offending, and predicting an individual's risk of causing harm whilst in custody or in the community.[5]

8. Indeed, as the Ministry of Justice acknowledges in its recently published AI action plan for justice:

   "Non-generative AI, including statistical models, deep learning, and other predictive machine learning methods, remains a cornerstone of the justice system's digital transformation."[6]

---

[1] Foxglove. (2025, June 26). *We forced the DWP to explain its benefits fraud algorithm: here's what we found*. https://www.foxglove.org.uk/2025/05/06/dwp-explain-benefits-fraud-algorithm/

[2] Department for Science, Innovation and Technology. (2024, February 12). *Introduction to AI assurance*. GOV.UK. https://www.gov.uk/government/publications/introduction-to-ai-assurance/introduction-to-ai-assurance#ai-assurance-in-context

[3] Edgar, K. (2018). *Ensuring a safe environment: a conflict centred strategy*. Prison Reform Trust. https://prisonreformtrust.org.uk/publication/ensuring-a-safe-environment-a-conflict-centred-strategy/

[4] For more information on categorisation visit our advice guide https://prisonreformtrust.org.uk/adviceguide/categorisation-mens-prisons/

[5] Zilka, M., Sargeant, H., & Weller, A. (2022). Transparency, Governance and Regulation of Algorithmic Tools Deployed in the Criminal Justice System: a UK Case Study. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (AIES '22)*, 880–889. Association for Computing Machinery. https://doi.org/10.1145/3514094.3534200

[6] Ministry of Justice. (2025). *AI action plan for justice*. https://www.gov.uk/government/publications/ai-action-plan-for-justice/ai-action-plan-for-justice

Examples of tools currently in use

9. The Offender Assessment System (OASys) is a structured professional judgement (SPJ) tool for staff working in prison and probation services. It incorporates some algorithmic instruments but is primarily designed to structure professional thinking in coming to a judgement or rating. It generates several predictive scores based on static (unchanging) and dynamic (changeable) risk factors.[7] Key components include:

    - **Offender Group Reconviction Score (OGRS)** — This predicts the likelihood of proven reoffending within one or two years of release from prison or the termination of a non-custodial sentence. It uses static factors such as age, gender, and criminal history.
    - **Risk of Serious Recidivism (RSR)** — This tool predicts the likelihood of an individual committing a serious harmful reoffence within two years. It includes sub-scores for contact sexual reoffending, indecent image reoffending, and non-sexual violence.
    - **Risk of Serious Harm (RoSH)** — This assesses the likelihood of life-threatening or traumatic events. It categorises risk levels as Low, Medium, High, or Very High, with separate scores for risks to children, known adults, the public, staff, and other prisoners.

10. From what is publicly known, the development of AI tools within prison settings appear to be replicating and building on existing algorithmic tools which have been in use for several years. These include:

    - **Digital Categorisation Service (DCS)** — an algorithm to assist staff in determining the security categorisation of a prisoner. This influences decision making when setting the level of physical security for a prisoner and their progression through the prison system. DCS draws on unspecified law enforcement data from a person's journey through the criminal justice system to produce a recommended initial categorisation, which can then be confirmed or overridden by the member of staff responsible for the process.[8]
    - **Violence in Prisons Estimator** — an algorithmic tool to estimate the number of violent incidents a prisoner is at risk of being involved in over the next year, based on age; history of violent incidents in custody; assault rates (overall and within the last 12 months); and time spent in custody. This is designed to assist prison staff in managing violence within prisons by identifying individuals who may require additional support or monitoring. The tool's estimates are also included as part of the DCS and has been widely adopted. Approximately 2,000 users access it monthly, and it supports around 50,000 categorisation decisions and 83,000 recategorisation decisions annually.[9]

---

[7] Zilka, M., Sargeant, H., & Weller, A. (2022). Transparency, Governance and Regulation of Algorithmic Tools Deployed in the Criminal Justice System: a UK Case Study. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (AIES '22)*, 880–889. Association for Computing Machinery. https://doi.org/10.1145/3514094.3534200

[8] Ministry of Justice. (2025, April 28). *Security categorisation policy framework*. GOV.UK. https://www.gov.uk/government/publications/security-categorisation-policy-framework

[9] Ministry of Justice. (2025, April 28). *HMPPS: Violence in Prisons Estimator*. GOV.UK. https://www.gov.uk/algorithmic-transparency-records/hmpps-violence-in-prisons-estimator

Future plans

11. The Ministry of Justice's AI Action Plan for Justice sets out its clear ambition to continue to further integrate AI into justice services in the next three years.[10] Cited examples include:

    - **Productivity Tools** — Deploying AI assistants to reduce administrative burdens, such as transcription tools for probation officers and semantic search tools for faster decision-making.
    - **Risk Assessment Models** — Using predictive tools to estimate violence risk in custody and support better decision-making.
    - **Personalised Rehabilitation** — Exploring AI-powered learning platforms to tailor education and rehabilitation for people in contact with the criminal justice system.

Questions about the use of algorithmic tools

12. On first inspection, that many of the tools we are currently aware of appear to automate existing non-generative algorithmic processes, may provide reassurance. The DCS tool for example, increases the likelihood of a more consistent application of the criteria that determine categorisation. In digitising the process, it also opens the possibility of regular and more sophisticated monitoring, both locally and centrally. However, there are a number of areas which we think could benefit from the committee's scrutiny.

Data accuracy

13. The effectiveness of any algorithm will always depend on the quality of data it is provided with and trained on.

14. As Zilka et al highlight:

    "Although data infrastructure systems do not usually include complicated algorithms or artificial intelligence, the quality and consistency of the data collected are essential prerequisites for reliable analytical analysis or predictions. It is, therefore, crucial to ensure that the data collection process achieves good quality, fit for purpose data and does not unnecessarily infringe on individual privacy."11

15. The Prison Reform Trust runs an advice and information service for prisoners and their families. Through that service as well as our other work in prisons, we are aware of examples where prisons are currently struggling to address the integrity of their data, and prisoners face challenges to get inaccurate or out of date information removed from their OASys report.

16. This can follow a prisoner throughout their sentence, with Prison Offender Managers copying and pasting inaccurate information across multiple OASys reports — making it difficult to trace and correct the original error. If AI is used carelessly in

[10] Ministry of Justice. (2025). *AI action plan for justice*. https://www.gov.uk/government/publications/ai-action-plan-for-justice/ai-action-plan-for-justice

[11] Zilka, M., Sargeant, H., & Weller, A. (2022). Transparency, Governance and Regulation of Algorithmic Tools Deployed in the Criminal Justice System: a UK Case Study. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (AIES '22)*, 880–889. Association for Computing Machinery. https://doi.org/10.1145/3514094.3534200

administrative tasks, such as those suggested by the Ministry of Justice in the 'Future plans' section above, there is a risk that the deployment of AI tools could accelerate the generation of inaccurate or misinterpreted information through a feedback loop, introducing bias; undermining decision making; and impacting the treatment and conditions of individuals in prison.

17. The Prison Reform Trust is also undertaking a research study as part of the Ministry of Justice Data First programme[12], which makes a large administrative dataset on criminal justice outcomes available to researchers. It has become apparent from this work that data can often be recorded inaccurately at source, such as the type of sentence people are serving and the terms under which they were released.

18. These variables, if used for modelling reoffending rates, are highly sensitive and consequential for sentencing policy. We suspect that this occurs owing to complex sentencing and release rules where nuances are not always understood by staff processing prisoners on the ground. Such reliability issues are still being explored and uncovered by Data First, which is actively seeking feedback on the quality of its dataset. But in the meantime, the development of algorithms continues.

Discrimination and bias

19. Even when data are accurately recorded there is a risk that the undoubted benefits of consistency mask the entrenchment of systematic unfairness.

20. BAME and Muslim men are disproportionately represented in the prison population. But that disproportionality is even more marked in their representation in higher security categories.

21. A careful and admirably detailed Equalities Impact Assessment carried out by the prison service for the Security Categorisation Policy Framework examined this phenomenon and found that the DCS tool did not appear to be the cause of a small increase in the disproportionality of outcomes (once factors such as the location of the pilot prisons had been considered), **but nor did it reduce it**.[13]

22. Zilka et al argue that this is poor proxy for assessing a tool's performance. "The baseline often used in this context is whether using the tool increases disproportionate outcomes…even when disproportionately does not increase following the adoption of a tool, it does not necessarily mean the tool is unbiased."[14]

23. The unanswered question, therefore, is whether disproportionality is being driven by decisions and information from processes that precede the use of the DCS tool and the data on which it depends. Specifically, intelligence about BAME and Muslim men which cannot be effectively challenged, and which may reflect conscious or unconscious bias.

---

[12] Ministry of Justice. (2025, July 31). *Ministry of Justice: Data first*. GOV.UK. https://www.gov.uk/guidance/ministry-of-justice-data-first

[13] Ministry of Justice & HM Prison and Probation Service. (2019). *Security Categorisation - Equalities Impact Assessment*. WhatDoTheyKnow. Retrieved September 2, 2025, from https://www.whatdotheyknow.com/request/use_of_digital_risk_assessment_t/response/1589030/attach/8/Equality%20Analysis.pdf

[14] Zilka, M., Sargeant, H., & Weller, A. (2022). Transparency, Governance and Regulation of Algorithmic Tools Deployed in the Criminal Justice System: a UK Case Study. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (AIES '22)*, 880–889. Association for Computing Machinery. https://doi.org/10.1145/3514094.3534200

24. Turning to the Violence in Prisons Estimator, the accompanying algorithmic transparency record suggests that the tool was evaluated "using a comprehensive approach, considering its performance, interpretability, and ethical implications" and that this involved "a thorough model bias evaluation and assessment of equality impacts".[15] Whilst this is a helpful indication of the seriousness the department places on the removal of bias from its tool, readers are being asked to take this judgement on face value, with no accompanying information of the analysis or its outputs.

25. Whilst there is certainly scope for improvement, such analysis and assessment will continue to be vital, acting as an important check on the development of effective policy, and avoiding unintended consequences, or exacerbating inequalities. It is vital that government can demonstrate, in court if needs be, their thinking and decision making in the development of AI tools.

Interpretability and explainability

26. It is tempting to simply exclude certain protected characteristics as a shortcut to "remove" bias within the tool. However, as previous examples have shown, this fails to determine whether a tool is unbiased or not.[16] Furthermore, their exclusion can perversely increase the risk of disproportionality in outcomes between groups through the emergence of proxy variables.[17]

27. Proxy variables are data points that indirectly represent or correlate with protected characteristics (such as race, gender, or religion) even when those characteristics are explicitly excluded from a dataset. These variables can perpetuate bias in algorithmic tools because they allow the system to infer or approximate the protected characteristic, leading to discriminatory outcomes. But they also undermine efforts to effectively interrogate the outcomes that tools produce — particularly if staff are inadequately equipped to do so, something which is exacerbated by the current high levels of prison officer turnover.[18]

28. The scientific nature of algorithmic tools can be seen as impartial and credible in a way that human judgements are not.[19] However, evidence suggests that algorithmic

[15] Ministry of Justice. (2025, April 28). *HMPPS: Violence in Prisons Estimator*. GOV.UK. https://www.gov.uk/algorithmic-transparency-records/hmpps-violence-in-prisons-estimator

[16] Zilka, M., Sargeant, H., & Weller, A. (2022). Transparency, Governance and Regulation of Algorithmic Tools Deployed in the Criminal Justice System: a UK Case Study. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (AIES '22)*, 880–889. Association for Computing Machinery. https://doi.org/10.1145/3514094.3534200

[17] Simons, J., Bhatti, S. A., & Weller, A. (2021). Machine Learning and the Meaning of Equal Treatment. *AIES '21: Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 956–966. https://doi.org/10.1145/3461702.3462556

[18] Ministry of Justice. (2025). *HM Prison and Probation Service workforce quarterly: June 2025.* GOV.UK. https://www.gov.uk/government/statistics/hm-prison-and-probation-service-workforce-quarterly-june-2025/hm-prison-and-probation-service-workforce-quarterly-june-2025

[19] Werth, R. (2018). Theorizing the Performative Effects of Penal Risk Technologies: (Re)producing the Subject Who Must Be Dangerous. *Social & Legal Studies*, *28*(3), 327–348. https://doi.org/10.1177/0964663918773542

tools produce systematically biased outcomes for particular groups,[20] and it can take time for these to become apparent, or be revised. Risk assessment tools are inexorably tied to the social world and all its biases, and it is worth paying more attention to how this may apply in reaching decisions with the potential to affect multiple ECHR articles.

"AI interpretability focuses on understanding the inner workings of an AI model while AI explainability aims to provide reasons for the model's outputs."[21]

29. Both interpretability and explainability are crucial for staff to understand the tools they are using, and to ensure they can command procedural fairness and legitimacy amongst prisoners.

30. As the Security Categorisation Policy Framework states:

    "The Governor must nominate a manager whose responsibility it is to ensure that the categorisation/recategorisation process is functioning effectively; that decisions are fair, consistent and taken without bias; to provide quality assurance of decision making; to collect and analyse data in terms of protected characteristics (see paragraph 4.3) alongside other equalities data to ensure that there is a complete picture of any disproportionate impact, and to implement change where necessary."[22]

31. But there is no indication of how local managers are to be equipped to carry out such an analysis, or what changes they might be empowered to make once they have.

32. In practice, for some prisons, the limited quantity of data available will make it hard to draw conclusions, certainly in the short term. The same cannot be said for national data, however, and whilst the equality analysis for the new policy framework is good in many respects, it leaves unanswered the question of why BAME and Muslim prisoners seem to fare less well both on initial categorisation and in the rarer cases where people are re-categorised to a higher security level.

33. Our own work on the sentence progression of people serving long sentences found that prisoners are often confused and disillusioned by the apparently simple proposition that they are required to reduce 'risk'.[23] This catch-all term masks important details—risk of what, from what, to whom, in what circumstances? It also provides a lesson in the importance of clarity, interpretability and explainability.

34. Part of the confusion around risk reduction stems from a mismatch between what prisons appear to expect from prisoners—broadly, compliance with the rules—and what those in probation and the Parole Board are looking for prisoners to demonstrate to secure their own development and eventual release. Participants told us that this was leading to them spending years of "nothing time" in prison. Years,

---

[20] Eckhouse, L., Lum, K., Conti-Cook, C., & Ciccolini, J. (2018). Layers of Bias: a unified approach for understanding problems with risk assessment. *Criminal Justice and Behavior*, *46*(2), 185–209. https://doi.org/10.1177/0093854818811379

[21] McGrath, A., & Jonker, A. (2025, April 17). What is AI interpretability? *IMB*. https://www.ibm.com/think/topics/interpretability

[22] Ministry of Justice. (2025, April 28). *Security categorisation policy framework*. GOV.UK. https://www.gov.uk/government/publications/security-categorisation-policy-framework

[23] Jarman, B., & Vince, C. (2022). *Making Progress?: What progression means for people serving the longest sentences*. Prison Reform Trust. https://prisonreformtrust.org.uk/prison-system-failing-to-prepare-long-term-prisoners-for-release/

often in the middle part of their sentence, where the sentence felt purposeless and stagnant.

35. These individual experiences raise a fundamental question about the fairness of predictive modelling: is a person's similarity to others in a dataset a fair basis for deciding on where they should live within the prison system, how they are managed, and whether they should progress? Current systems do theoretically allow for balancing algorithmic information with professional judgement about the individual, and it is important that the development of more advanced AI tools does not result in an even more opaque decision-making system for prisoners.

Future development

36. The project lifecycle model outlined in the Ministry of Justice's AI and data science ethics framework mirrors many similar approaches, with its design, development, and deployment stages.[24]

37. The Violence in Prisons Estimator, uses the protected characteristic of age as a key metric to inform a person's violence risk score.

38. It is especially important when decisions are taken which have implications for an individual's Article 5 rights, that there are sufficient safeguards to prevent the conflation of correlation and causation when determining which characteristics reliably predict outcomes.

39. A rapid evidence assessment completed for the Ministry of Justice in 2018 found:

    "The most firmly supported variable is younger age, and there is moderately high support for previous involvement in violence while in prison, for having a conviction for a violent offence, a history of drug offences or drug abuse, and for gang involvement. Evidence for other variables including educational level and race/ethnicity is more mixed."[25]

40. It would be helpful to understand from officials what safeguards are in place to ensure that these are sufficiently interrogated to determine causation, before integration into the model if other protected characteristics are seen to correlate with a higher risk of violence in future.

Prison conditions and the prevention of inhumane and degrading treatment

41. Imprisonment today is far more punitive than just the deprivation of liberty. Living conditions in many prisons, including newer facilities, have deteriorated due to overcrowding and neglect.[26] Such conditions can be experienced as even more

[24] Ministry of Justice. (2025). *Ministry of Justice AI and Data Science Ethics Framework: Introduction to the SAFE-D framework*. GOV.UK. https://www.gov.uk/government/publications/ministry-of-justice-ai-and-data-science-ethics-framework

[25] McGuire, J. (2018). *Understanding prison violence: a rapid evidence assessment*. Ministry of Justice. https://www.gov.uk/government/publications/understanding-prison-violence-a-rapid-evidence-assessment

[26] HM Chief Inspector of Prisons. (2024). *Annual Report 2023-24* (HC 218). House of Commons. https://hmiprisons.justiceinspectorates.gov.uk/hmipris_reports/annual-report/

punitive when a person is detained in custody for preventative reasons of public protection — and for some — prison is a wholly unsuitable and inappropriate environment.

42. In 2024, the National Preventive Mechanism reported that rising prisoner numbers, overcrowding, staff shortages, and decaying infrastructure have created untenable conditions. [27]

43. “We have seen buildings that are not fit for purpose brought back into use, or two people being squeezed into single cells, to cope with more prisoners, without enough staff or wider infrastructure to adequately prepare them for release and manage risks of reoffending.”

44. Inspectors have found cells so crowded that that beds are placed directly next to toilets with little or no screening, and shared areas are unhygienic and unsafe. Filthy showers, broken heating systems, inadequate ventilation, and infestations of rats, cockroaches and even pigeons reflect the chronic underinvestment in prison maintenance.[28]

45. Mental health support is equally inadequate. With long waits for forensic mental health beds, severely unwell individuals are left in settings ill-equipped for their needs. This leaves staff struggling, and prisoners at risk of irreversible harm. Rates of self-harm have reached record highs, with 76,365 incidents reported in the 12 months to June 2024—equating to 876 incidents per 1,000 prisoners.[29]

46. Prisons are violent places: in the 12 months to March 2024, the rate of assaults was 327 assaults per 1,000 prisoners (28,292 assaults), up 19% from the 12 months to March 2023.[30]

47. Older prisoners face unique challenges in these crumbling systems. The number of prisoners aged 60 or older has surged by 82% in the last decade and by 243% since 2002. With higher prevalence of chronic illnesses, disabilities, and mobility issues, prisons remain poorly equipped for their care. Many spend excessive time confined to cells, exacerbating isolation and hindering rehabilitation.[31]

48. Women in prison often have histories of trauma, substance misuse, and mental health difficulties, with many serving sentences linked to offences committed in coercive or abusive relationships. Despite making up less than 5% of the prison population, women account for disproportionately high rates of self-harm and suicide,

---

[27] National Preventive Mechanism. (2024). *Monitoring places of detention: 15th Annual Report of the United Kingdom’s National Preventive Mechanism 2023/24* (CP 1290). His Majesty’s Stationery Office. https://nationalpreventivemechanism.org.uk/document/fifteenth-annual-report-2023-24/

[28] HM Chief Inspector of Prisons. (2024). *Annual Report 2023-24* (HC 218). House of Commons. https://hmiprisons.justiceinspectorates.gov.uk/hmipris_reports/annual-report/

[29] Ministry of Justice. (2024). *Safety in custody: quarterly update to June 2024* [Dataset]. https://www.gov.uk/government/statistics/safety-in-custody-quarterly-update-to-june-2024

[30] Ministry of Justice. (2024). *Safety in custody: quarterly update to June 2024* [Dataset]. https://www.gov.uk/government/statistics/safety-in-custody-quarterly-update-to-june-2024

[31] House of Commons Justice Committee. (2020). *Ageing prison population* (HC 304). House of Commons. https://committees.parliament.uk/work/270/ageing-prison-population/publications/

now at record levels. The chaotic and punitive nature of prison life can exacerbate these vulnerabilities, with noisy and traumatic environments often mirroring past abuses.[32]

49. The current state of our prisons reflects a systemic failure to uphold basic human dignity. People are sent to prison as punishment, not for punishment. Yet poor conditions are adding an additional layer of punishment onto the deprivation of liberty. In their current condition prisons are places of harm rather than rehabilitation, perpetuating cycles of despair and reoffending.

50. Furthermore, despite the availability of high quality and consistent reporting on the poor state of conditions by independent scrutiny bodies, and the clear identification of where failures have occurred, the prison system appears unable to learn or adapt to the challenges it faces, undermined by the relentless demand placed upon it.[33]

51. All of this serves to contextualise the existing difficulties that the department has in meeting the most visible and obvious human rights obligations.

52. We recommend that the government revisits a previous attempt to strengthen the independent scrutiny bodies through legislation.[34] In our evidence to the 2020 consultation we highlighted our concerns about the effectiveness of independent scrutiny posed by inadequate resourcing; a failure on the part of ministers to respond promptly in all circumstances when serious criticisms are made; and the failure of operational services effectively to implement recommendations which they choose to accept.[35]

53. As the development and deployment of algorithmic and AI tools continues it will be critical that independent scrutiny bodies, particularly HM Inspectorate of Prisons, HM Inspectorate of Probation, and the Prisons and Probation Ombudsman are able to effectively strengthen their scrutiny of the use of algorithmic and AI tools.

54. We are concerned that there is a current deficit in the scrutiny provided to psychology and Offender Management Unit work within prisons, with too much focus

---

[32] Kelman, J., Gribble, R., Harvey, J., Palmer, L., & MacManus, D. (2022). How does a history of trauma affect the experience of imprisonment for individuals in women's prisons: a Qualitative exploration. *Women & Criminal Justice*, *34*(3), 171–191. https://doi.org/10.1080/08974454.2022.2071376

[33] See Prisons Watchdog has "no teeth." (2021, July 19). *Insidetime*. https://insidetime.org/newsround/prisons-watchdog-has-no-teeth/;
Inquest. (2020). *Deaths in prison: a national scandal*. https://www.inquest.org.uk/deaths-in-prison-a-national-scandal; and
House of Commons Justice Committee. (2018, February 16). *Boost Inspectorate to reverse decline in prisons* [Press release]. Retrieved September 4, 2025, from https://committees.parliament.uk/committee/102/justice-committee/news/99536/boost-inspectorate-to-reverse-

[34] Ministry of Justice. (2020). *Strengthening the Independent Scrutiny Bodies through Legislation*. GOV.UK. https://consult.justice.gov.uk/digital-communications/strengthening-the-independent-scrutiny-bodies/

[35] Prison Reform Trust. (2020). *Prison Reform Trust response to the Ministry of Justice consultation on strengthening the independent scrutiny bodies through legislation*. https://prisonreformtrust.org.uk/consultationresponse/strengthening-the-independent-scrutiny-bodies-through-legislation/

on whether a process "was done" rather than "how was it done".[36] Such an approach risks limiting scrutiny outside of the black box, where only those in HMPPS are able to assess performance on risk matters.

Recommendations

55. Given its extensive experience with non-generative algorithmic tools, its ambition to lead on AI adoption and regulation in government, and the potential for its tools to engage multiple ECHR articles, we recommend the committee invite ministers or officials from the department to give oral evidence.

56. We also suggest the following requirements for the future operation of this technology:

    - People should be informed about the presence of algorithmic tools when they contribute to a decision making process, and be provided with information about the tool in simple and non-technical language. This is particularly important given the low levels of literacy, and digital literacy amongst sections of the prison population.
    - Tools must transparent and interpretable by the staff who use them.
    - Staff should be trained to effectively scrutinise and monitor the results they produce, with clear national guidance setting out how this should be achieved.
    - National guidance should be published for prison governors on how to interpret their local data to accurately identify outcomes that are disproportionate and which require explanation or reform, in line with the recommendation of the Lammy Review.[37]
    - People must be able to effectively challenge a decision, and the information used to inform a decision; there must be the possibility of a binding requirement from an external and independent arbiter to remove data when found to be inaccurate, with a subsequent full review of the decision.
    - Regular national publication of the data the system produces, and a refusal to accept disproportionate outcomes which cannot be adequately explained. This should also include an assessment on the use of tools by staff and whether there is any change in attitudes to decision-making and outcomes.
    - The government should strengthen independent scrutiny bodies through legislation.
    - HM Inspectorate of Prisons, HM Inspectorate of Probation, and the Prisons and Probation Ombudsman should be equipped to provide effective scrutiny of the use of algorithmic and AI tools.

[36] See Prison Reform Trust. (2023, April 19). *Offender Management and sentence planning | Prison Reform Trust*. https://prisonreformtrust.org.uk/adviceguide/offender-management-and-sentence-planning/ for more information

[37] Lammy, D. (2017). *Lammy review: final report*. GOV.UK. https://www.gov.uk/government/publications/lammy-review-final-report