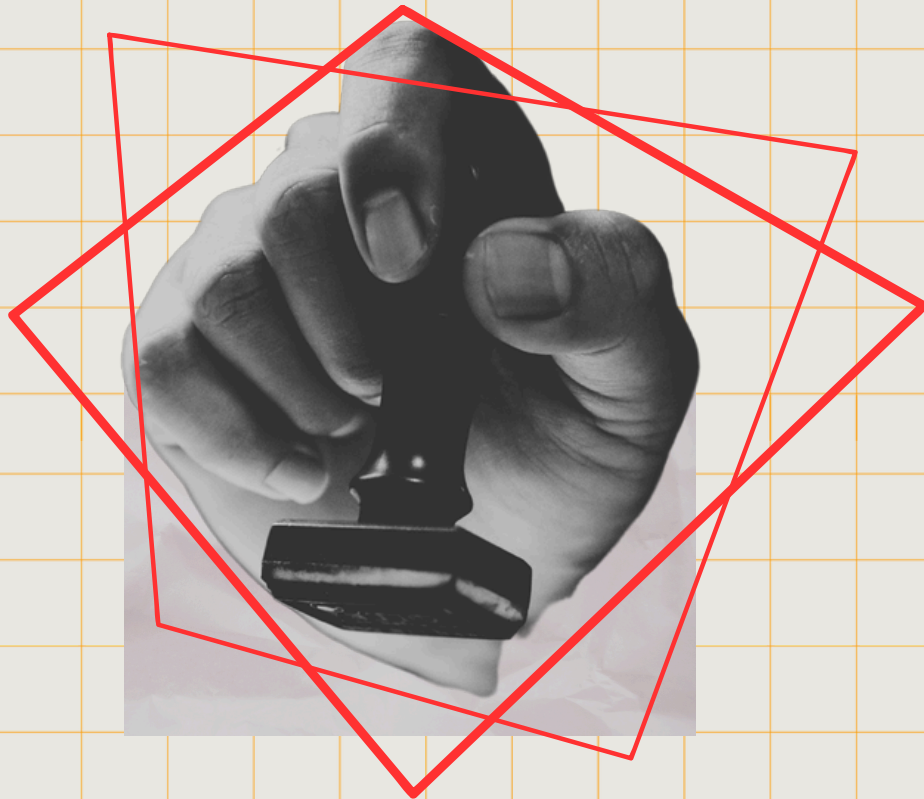


JULY 2025

ALGORITHMIC ACCOUNTABILITY IN THE UK

HOW FOIA SHEDS LIGHT ON AUTOMATED WELFARE
BY MORGAN CURRIE & ALLI SPRING



Algorithmic Accountability in the UK

How FOIA Sheds Light on Automated Welfare

Authors: Dr Morgan Currie, Dr Alli Spring

The University of Edinburgh

CC BY 4.0

How to cite this paper: Currie, M., Spring, A. (2025) 'Algorithmic Accountability in the UK - How FOIA Sheds Light on Automated Welfare', The University of Edinburgh, Edinburgh, available at: <http://dx.doi.org/10.7488/era/6168>

If you require this document in an alternative format, such as large print or a coloured background, please contact Morgan Currie, morgan.currie@ed.ac.uk.

Acknowledgements:

The authors would like to thank Lena Podoletz for her help collating and analysing several of the FOI requests in our dataset. Thanks to Jake Hurfurt for his ongoing support of this project. Thanks as well to MySociety for supporting research into algorithmic accountability using WhatDoTheyKnow.

Cover art: Max Perry

Executive Summary 3

Background 5

Data-intensive Fraud Detection in Social Security 5

About the Freedom of Information Act 6

The Good Faith Mosaic Method 7

Findings 9

RIS and IRIS Phase, 2018-2022 9

The Advances Model, 2022-2023 14

Future Modelling 17

Conclusion21

References 22

Tables 26

Glossary 30

Executive Summary

For the past several years, the UK's Department of Work and Pensions' (DWP) has adopted automated data analytics and machine learning to detect fraud and error in social security applications. The department has disclosed very little about these initiatives to the public beyond cursory mentions in year-end accounting reports and answers to parliamentary questions, despite growing interest from politicians, grassroots groups and citizens in how they affect welfare applicants.

One way to gain a larger picture of the DWP's practices is to look at the results of Freedom of Information (FOI) requests people have submitted over the years to ask about DWP's data-intensive systems for fraud and error detection. This study analyses 51 such requests from 2018 up to the present (Table I); these present a chronology of some of the tools in place, the internal processes around them, how they are evaluated and how they relate to staff roles (see the glossary at the end of the report for a list of some of these services, tools and staff teams). The exercise also yielded a collection of 44 internal documents that DWP released to requesters (Table II).

We learn from the requests that DWP has piloted or uses two main automated approaches for fraud and error detection: 1) **data matching** of existing claimant records to records held by other public agencies, and 2) **risk scoring** of claims through supervised machine-learning models. With both processes, flagged claims go to at least one human caseworker for additional review.

To carry out the first approach, DWP has created data sharing agreements with several other government agencies to identify housing, student, mortgage and immigration fraud and error. It also has agreements with private sector companies, including the credit referencing agency Indessex to help identify claimants with an undeclared partner and with banks to detect capital and abroad fraud and error.

To carry out the second approach, DWP has developed at least five models for risk scoring claimants: 1) an Advances Model to risk score people applying for a Universal Credit loan or advance during the first month of being approved for the benefit; 2) the Living Together Model, 3) the Self-employed Earnings Model, 4) the Housing Model, and 5) a model to detect undeclared capital (the last four are in pilot phase at the time of the requests). We get some insight into DWP's evaluations of its modelling through the release of an Equality Analysis of the Advances Model, which found bias in the model against protected characteristics of age, disability, marital status and nationality.

Requests also yield some details about DWP's internal structure: as of a 2023 request, we find that a Risk Review Team uses data matching (but not modelling) on claims already in payment, while an

Enhanced Checking Service checks new claims; the exact tools these teams use are not disclosed. Both are part of the Enhanced Review Team. A separate Open Source team investigates social media data of claimants suspected of committing fraud. A process called Targeted Case Reviews entails checking a sample of automatically flagged in-payment claims for error.

But the vision remains partial. DWP only successfully answered seven out of 51 requests, and 24 were only partially answered (see Table III). We still lack understanding of all the teams in place that work with these tools, how many of these tools and systems there are, and how they are evaluated. We also lack a clear definition of terms, with technical terms such as 'data matching,' 'data analytics,' 'machine learning,' 'risk modelling,' and 'Artificial Intelligence' (AI) – which are used throughout the DWP FOI requests, responses and released internal documents – often occurring interchangeably and without technical specification.

Particularly concerning, we find that DWP takes a non-standard approach to answering FOI requests. Requesters who take the time and have the knowledge of legal FOIA procedures achieved much greater success than other requesters asking for the same information. For four requests, DWP released documents after a complaint to the ICO – as a result, requesters seeking the same information but not taking this onerous step did not receive these documents.

Overall, this analysis shows that there is a need for the DWP to be more forthcoming to the public about its automated fraud and error detection tools – including disclosing information about the internal safeguards deployed to protect claimants from their potential risks – and to be more consistent in releasing information about these technologies.

Automated Fraud Detection in Social Security

Over the last decade, governments worldwide have been quietly adopting data analytics, data matching and predictive risk-profiling tools to counter fraud and error in social security benefits.¹ Civil society has reacted by sounding the alarm that these techniques put vulnerable populations under increased surveillance and can lack transparency in how decisions are made.² The Dutch System Risk Indication (SyRI), for instance, collated data from several government agencies to find anomalies that might signal welfare fraud, but it was ruled unlawful by a Dutch court for over-targeting immigrants.³

In the UK, the Department of Work and Pensions has subjected social security claims to data matching and data analytics since at least 2010, with hopes to recoup billions of pounds of public funds lost to fraud.⁴ Since 2022 the Department has deployed machine learning models to risk profile cases that may be fraudulent or in error,⁵ flagging high-risk cases for further questioning by staff. The DWP takes steps to mitigate potentially negative consequences of these tools through fairness testing and by always having a human in the decision-making loop.⁶

Still very little is known about the DWP's use of data analytics and modelling to deter fraud. The Committee of Public Accounts has called for DWP to assess any bias in machine learning models and any impact on specific groups.⁷ The UK Information Commissioner has warned the DWP it risks being in contempt of court for not releasing details of its modeling techniques.⁸ In civil society, journalists and

¹ Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Publishing Group; Elyounes, D. (2021). "Computer Says No!": The Impact of Automation on the Discretionary Power of Public Officers. *Vanderbilt Journal of Entertainment & Technology Law*, 23(3), 451; Dencik, L. (2022). The Datafied Welfare State: A Perspective from the UK. In A. Hepp, J. Jarke, & L. Kramp (Eds.), *New Perspectives in Critical Data Studies: The Ambivalences of Data Power* (pp. 145–165). Springer International Publishing. https://doi.org/10.1007/978-3-030-96180-0_7.

² Big Brother Watch. (2021, July 20). *Poverty panopticon. The hidden algorithms shaping Britain's welfare state*. <https://bigbrotherwatch.org.uk/wp-content/uploads/2021/07/Poverty-Panopticon.pdf>, Burgess et al. 2023; Burgess, M. (n.d.). *This Algorithm Could Ruin Your Life*. *Wired*. Retrieved 21 June 2025, from <https://www.wired.com/story/welfare-algorithms-discrimination/>

³ van Bekkum, M., & Borgesius, F. Z. (2021). Digital welfare fraud detection and the Dutch SyRI judgment. *European Journal of Social Security*, 13882627211031257. <https://doi.org/10.1177/13882627211031257>

⁴ <https://www.nao.org.uk/wp-content/uploads/2023/07/dwp-report-on-accounts-2022-23.pdf>

⁵ <https://assets.publishing.service.gov.uk/media/62d52f55e90e071e814807fd/annual-report-accounts-2021-22-web-ready.pdf>

⁶ <https://www.nao.org.uk/wp-content/uploads/2024/07/dwp-report-on-accounts.pdf>

⁷ *Ibid.*

⁸ <https://www.theguardian.com/politics/2023/sep/03/uk-warned-over-lack-transparency-use-ai-vet-welfare-claims>

activists are calling for disclosures on the effectiveness of these approaches or if they pose concerns for profiling certain populations. Charities such as the Child Poverty Law Centre, Public Law Project and Big Brother Watch have applied Freedom of Information requests, written evidence to public reviews, carried out audits and led campaigns to create more due process and transparency around DWP's practices.⁹

About the Freedom of Information Act

This report adds to these efforts by analysing the results of Freedom of Information requests submitted by individuals and charities to gain information about DWP's automated fraud detection activities. In the UK, the 2000 Freedom of Information Act (FOIA) gives citizens the right to access information held by the government and other public authorities, with agencies expected to respond to requests within 20 working days.¹⁰

There are two important features of the FOIA that bear on our analysis. The first are the many exemptions.¹¹ Several sections of the FOIA cite reasons that an agency can refuse a request, ranging from instances when the information is already accessible or will be in the future, to national security concerns and violations of commercial trade secrets. Agencies may also refuse a request if it is too onerous and costly, in which case, the agency can suggest the requester narrow their request.

For our dataset, we follow WhatDoTheyKnow's labels of 'refused,' 'successful,' 'partially successful' or 'information not held.' The DWP refuses requests when it supplies no answers due to FOIA exemptions. Successful requests are answered with no cited exemptions. The majority of requests cite exemptions for some parts of the request but not others, or redact documents based on exemptions – these are partially successful. Some requests are also not answered due to the agency's not holding the information ('information not held'). The majority of refused or partially refused requests cite the exemption Section 31, applied when releasing information would harm an agency's ability to enforce the law. DWP cites Section 31 out of concern that potential fraudsters would use the released information to undermine, or game, their fraud-detection systems.

A second feature of the FOIA is that requesters can challenge agencies if they are dissatisfied with the answer. There are two mechanisms to raise a challenge. The first is to ask for an internal review of the request, which leads to a second party within the agency reviewing the request, then determining if

⁹ See <https://publiclawproject.org.uk/content/uploads/2021/10/PLP-Judicial-Review-and-Courts-Bill-2nd-Reading-Briefing-clause-1-and-2.pdf>, <https://bigbrotherwatch.org.uk/campaigns/welfare-data-watch/> and <https://cpag.org.uk/what-we-do/project-work/projects-england-and-wales/universal-credit-digitalisation-and-rule-law>

¹⁰ <https://ico.org.uk/for-organisations/foi/>

¹¹ <https://ico.org.uk/for-organisations/foi/guide-to-managing-an-foi-request/exemptions/list-of-exemptions/>

the answer has a sound basis, legal or otherwise.¹² If the requester remains dissatisfied, they can make a complaint to the UK Information Commissioners Office (ICO), triggering an external audit of the process by the ICO.¹³

Several of the requests we analysed entailed a lengthy back and forth with the DWP, especially when requesters asked for an internal review, then ultimately challenged the DWP's responses with an ICO complaint.

The Good Faith Mosaic Method

To create the corpus of FOI requests we used the website WhatDoTheyKnow,¹⁴ run by UK civic tech organisation MySociety. This platform allows people to submit requests to organisations and archives those results for public viewing and download.

Using the website's search function, we made the following keyword queries: 'fraud DWP algorithm,' 'fraud dwp automation,' 'fraud dwp machine learning,' 'fraud dwp risk assessment,' 'fraud dwp AI,' 'fraud dwp artificial intelligence,' 'fraud dwp AI risk score,' 'fraud dwp model,' 'fraud DWP automation' and 'fraud dwp machine learning.' In all cases, our search added the 'requested_from:dwp' term to limit results to those answered by the DWP.

The site offers the action to 'download a .zip file of all correspondence.' Using this function, we downloaded all correspondence between the requester and the DWP, plus any primary source documents released through the request. We organised the requests chronologically and gave each a unique ID number (R1, R2, etc). We collected 44 primary documents released by the DWP as part of our dataset (and three differently redacted versions of one of these documents).

Our analysis also includes all public documents and media reports referred to in the FOI requests.

This approach of amassing public FOI requests helps us better understand how the DWP uses data matching, data analytics, and machine learning models to detect fraud and error in benefit applications. This method also tells us how the DWP discloses information to the public – specifically patterns in responses and response timeframes. Our analysis operates at two levels of inquiry: first, of the state of DWP's data-intensive fraud detection – as far as we can know it – through the patchwork of document releases; second, of the interplay between requesters using legal manoeuvres, such as ICO complaints, to gain this information and the DWP's use of lawful FOI exemptions to keep this information at bay.

¹² <https://ico.org.uk/for-organisations/foi/guide-to-managing-an-foi-request/complaints-internal-reviews/>

¹³ <https://ico.org.uk/make-a-complaint/foi-and-eir-complaints/foi-and-eir-complaints/>

¹⁴ <https://www.whatdotheyknow.com>

The 'mosaic effect' is a term used to describe a route to harm from FOI disclosures.¹⁵ This effect can occur when disclosed information, even if not of concern on its own, is combined with other public information in a way that leads to harm or prejudice. We consider our combining a collection of FOI requests together along with other public documents a '**good faith mosaic**' method, as we piece together a picture of DWP's practices for public benefit and without intent to harm. The result is a gestalt of what we currently know, what remains undisclosed, and the tactics of civil society to understand as much as possible about DWP's practices in applying automation to fraud detection.

The study took part in March – July 2025. For data protection purposes, we redacted names from our research dataset and do not supply the names of requesters. We also acknowledge the limitations of our dataset: it does not give a full account of all FOI requests submitted to the DWP during the period under review but only those submitted through the WhatDoTheyKnow platform.

¹⁵ <https://ico.org.uk/for-organisations/foi/freedom-of-information-and-environmental-information-regulations/information-in-the-public-domain/#whatareprejudice>

Findings

In our collection, FOI requests typically follow the release of public documents that have offered some insight into a new data-intensive service or system – sometimes just a sentence or paragraph buried in the larger text. For activists and curious citizens, this small piece of information is like a slightly cracked, jammed door – people use FOI as a legal crowbar to get a more substantial look at a system or service of interest.

We organise our set of FOI requests into three different chronological phases, each largely corresponding to the release of public documents from the DWP and, less frequently, the National Audit Office (NAO), media articles or parliamentary records. The starring document type is the DWP's yearly *Annual Report and Accounts*, published every June, which has triggered the most requests.

The first phase, what we call the **RIS and IRIS Phase**, covers 2018-2022. The second phase, the **Advances Model Phase**, covers 2022-2023. The final phase, **Future Modelling**, begins in 2023 and brings us up to the present time of writing.

We go through each phase in the sections below, offering context before giving accounts of the type of FOI requests submitted and the answers that the DWP gives in turn.

RIS and IRIS Phase, 2018-2022

Context

This phase marks a period when requesters ask for details and documents about the DWP's data matching and risk analysis for fraud and error detection in benefit applications and claims. The majority of FOI requests¹⁶ relate to the publication of the DWP's *Annual Report and Accounts 2017-2018*. This report mentions that the DWP's Analysis and Intelligence (A&I) Hub is using 'new analytical tools' for fraud investigation and 'external datasets to risk assess' Universal Credit claims. The report also

¹⁶ Requesters also cite media articles, such as a Sun article on 'robots' that 'scour Facebook' for illegal activity: <https://www.thesun.co.uk/news/5238537/robots-will-scour-facebook-for-criminal-activity-in-a-2billion-benefit-fraud-blitz/#comments> and 2018 Forbes article <https://www.forbes.com/sites/bernardmarr/2018/10/29/how-the-uk-government-uses-artificial-intelligence-to-identify-welfare-and-state-benefits-fraud/>

mentions 'cutting-edge artificial intelligence and trials using algorithms'.¹⁷ These algorithms detect activity 'commonly used by fraudsters' by 'searching for anomalies in billions of items of data.' The document says the DWP will introduce a 'risk and intelligence service', or RIS, to 'improve our understanding of claimant behaviour' through data processing.¹⁸

Two years later the DWP's *Annual Report and Accounts 2019-2020* mentions a new Integrated Risk and Intelligence Service, or IRIS, that launched in March 2020.¹⁹ IRIS combines RIS with 'other intelligence teams' to prevent fraud using 'new data matching rules' that 'route cases to our investigators where discrepancies are identified.'²⁰ IRIS matches data from across government and third parties, including credit agencies.

Investigating some of these discrepancies is the DWP's Risk Review Team, or RRT.²¹ In a Parliamentary debate on 26 January, 2022, MP Kate Osamor described 29 constituents, many of whom are single mothers from Bulgaria, as having had their benefits suspended pending RRT reviews of their claims. 'Sadly,' she says, 'there did not appear to be much information publicly available about the risk review team.'²²

¹⁷ <https://assets.publishing.service.gov.uk/media/5b364091e5274a0bc32fabf7/dwp-annual-report-and-accounts-2017-2018.pdf>. For more on the A&I Hub, see this 2016 letter from DWP and HMRC to the House of Commons: <https://www.parliament.uk/globalassets/documents/commons-committees/public-accounts/Correspondence/2015-20-Parliament/PAC-Response-final-signed-copy-of-follow-up-letter-to-3rd-party-data.pdf>.

¹⁸ According to a 2016 letter to the Public Accounts Committee, this work dates to at least 2015, when DWP established the A&I Hub for Universal Credit. The A&I Hub uses 'data driven rules' 'to identify inconsistencies in benefit entitlement for Universal Credit claimants; it draws from third-party data, including that contained by Land Registry, Valuations Office, the National Fraud Initiative, Credit Reference Agencies and the Metropolitan Police. <https://www.parliament.uk/globalassets/documents/commons-committees/public-accounts/Correspondence/2015-20-Parliament/PAC-Response-final-signed-copy-of-follow-up-letter-to-3rd-party-data.pdf>. DWP's using data matching to detect fraud goes back to at least 2010, according to: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/214334/tackling-fraud-and-error.pdf

¹⁹ A June 2018 NAO report, *Rolling Out Universal Credit*, also mentions that DWP will launch the IRIS, "a fully automated risk analysis and intelligence system for fraud and error" that will finish in Spring 2019. <https://www.nao.org.uk/wp-content/uploads/2018/06/Rolling-out-Universal-Credit.pdf>.

²⁰ <https://assets.publishing.service.gov.uk/media/5ef9f3b2d3bf7f769392421e/dwp-annual-report-and-accounts-2019-2020.pdf>. The 2019-2020 *Annual Report and Accounts* describes the pandemic years' skyrocketing Universal Credit fraud, much of it by organised gangs; DWP responded to the issue with new data intensive approaches. See also a response from MP Will Quince, then Secretary of State for Work and Pensions, on how DWP is tackling fraud: <https://questions-statements.parliament.uk/written-questions/detail/2021-04-23/186257/>

²¹ See more in the UK Parliament's written question about RRT in 2021: <https://questions-statements.parliament.uk/written-questions/detail/2021-11-19/78475/>.

²² <https://hansard.parliament.uk/Commons/2022-01-26/debates/333BCD75-7B81-464A-BA13-91D711B1A4EF/DWPRiskReviewTeam>. Osamor also requested a cumulative total figure for claims suspended – at the time, it was 177,638, with 7,221 cases reinstated: <https://questions-statements.parliament.uk/written-questions/detail/2022-09-07/49224>.

FOI Requests

There were 21 requests during this period, starting in April 2018, with the last submitted May 2022.

Requesters during this time rely on various terms; these include 'robots,' 'artificial intelligence system,' 'artificial intelligence software,' 'AI based tools,' 'analytical tools' and 'software' to identify potential benefit fraud – all terms mirroring language in official reports or media. Requests also include inquiries about the A&I Hub, IRIS, the Risk Review Team and a DWP project called the General Matching Service.²³

Requesters want to know more about how IRIS and the A&I Hub use data and machine learning generally – one person requested, for example, 'any description of the Intelligent Automation Garage's robots used to detect benefit fraud' (R11). More granularly, people ask about the data sources in use, including from third parties; who designed the systems or services and whether third-party contractors were involved; and costs, both the costs of implementing the systems or services and any savings DWP has made by reducing fraud.

A large set of questions relate to oversight: people want to know how these initiatives are being evaluated for accuracy or for their impact on claimants, the numbers of people flagged, and whether anyone has requested Subject Access Requests about use of their data. Some requests ask for specific documents: privacy impact analyses, data protection impact analyses (DPIAs),²⁴ equality impact analyses,²⁵ data sharing agreements, memorandums of understanding, and application forms for data sharing. People also requested internal procedural documents, such as those outlining governance frameworks and strategies.

Four requests inquired into the DWP's use of AI to search social media.

²³ These FOI requests do not reference any particular information about the General Matching Service, so it is not known how the requesters knew of this service. The service has been mentioned in past reports, such as 2018-2022 Fraud, Error and Debt Strategy, Department for Work and Pensions:

<https://pdacounterfraud.co.uk/wp-content/uploads/2023/07/DWP-Fraud-Error-and-Debt-Strategy-2018-2022.pdf> and more recent information on the service is in the 15 May 2025 Quality statement: DWP benefits statistical summary: <https://www.gov.uk/government/publications/dwp-statistical-summary-policies-and-statements/quality-statement-dwp-benefits-statistical-summary>.

²⁴ Required under GDPR when an organisation embarks on a new activity involving the collection of personal data: <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/accountability-and-governance/data-protection-impact-assessments-dpias/what-is-a-dpia/>

²⁵ Assesses whether the policy has a disparate impact on persons with protected characteristics. See: <https://commonslibrary.parliament.uk/research-briefings/sn06591/>

DWP responses

Three of the 21 requests during this phase were successfully answered by DWP; these answered questions on the General Matching Service. We find out that the service ‘investigates compliance mismatches which could be as a result of customer or official error.’ Mismatches do not trigger a fraud investigation but can lead to a fraud investigation ‘carried out elsewhere.’ From another request, we find out the General Matching Service is not part of the National Fraud Initiative.²⁶

Nine requests were partially answered. Four of these yielded documents related to Universal Credit and the A&I Hub. Some of these documents give insight into DWP’s data sharing agreements with several other agencies, signed between 2016-2018. The DWP has data sharing agreements with: the Land Registry for property ownership data to determine mortgage-related fraud; the Driver and Vehicle Licensing Agency to determine housing composition-related fraud; the Cabinet Office and The Student Loans Company Limited within the National Fraud Initiative for student loan data, as students are ineligible for Universal Credit; the Home Office, to prevent payments to undocumented migrants; and HM Revenues and Customs for data on earnings and pension.²⁷ Other released documents relate to the DWP’s internal governance around data protection, including meeting minutes of the Data Protection Governance Board Meeting on three dates in 2019. All of these documents were lightly redacted per the Section 40 exemption, relating to personal data.

One request led to document releases after the requester asked for an internal review and made a complaint to the ICO (R18). After citing the FOIA Section 12 exemption – for when cost of compliance exceeds the appropriate cost limit – DWP released eight heavily redacted data protection impact analyses (DPIAs) related to IRIS. Six of these appear to describe data sharing between DWP and another party, while two (1442 and 1300) describe how DWP will apply risk modelling for fraud reduction. DPIA 1442 describes a model ‘trained on a selection of 1 million live [...] cases at prepayment stage.’ DPIA 1300 describes an Advances Model (more on this below) and a Common Risk Engine that entails risk scoring referrals ‘through to the Counter Fraud and Compliance Directorate’.

The remaining four partially answered requests did not include documents. For these, the DWP cites the FOIA Section 31 exemption, and, in two requests, Section 24, the national security exemption, for not disclosing more information, even when requesters ask for internal review.

²⁶ The National Fraud Initiative is “A data matching exercise operated by the Public Sector Fraud Authority that helps prevent and detect fraud”: <https://www.gov.uk/government/collections/national-fraud-initiative>

²⁷ These include privacy impact assessments from the ‘Living Together Data & Analytics’ project and the ‘Land Registry Property Ownership Data Pilot’; data sharing agreements with the Land Registry for property ownership data, with Minister of the Cabinet Office for student loan data, with The Student Loans Company Limited, and with the DVLA; DPIAs with Driver and Vehicle Licensing Agency for UC verification and the Student Loan Company; an Information Sharing Business Case for data sharing by the Student Loan Company; and DWP’s memorandums of understanding with the Land Registry, a Student Loan Company within the National Fraud Initiative, HMRC, and the Home Office.

From these requests, we learn more about DWP's internal processes. First, 'analytical tools' make referrals of risky claims to someone at the DWP to investigate so that 'no claimant is automatically denied benefit.' We also find out that the Risk Review Team (RRT) does not carry out fraud investigations but rather reviews UC claims based on additional evidence provided by claimants; any subsequently flagged applications go to specialist Fraud Investigators. We find out about contractors: the DWP has a contract with credit referencing agency Indessex to help identify claimants with an undeclared partner, and several companies have partnered with DWP over the years to develop its software to detect fraud (these are Oracle, Informatica, SAS, Cloudera, Amazon Web Services, Hewlett Packard Enterprise Services, BAE Systems Applied Intelligence, Kainos).

Among these successful and partially successful answers, the DWP will not answer general questions about third-party data providers or the datasets used by its A&I Hub or IRIS, nor provide staff training manuals or policies related to IRIS. The department will not release information on the accuracy and error rates or trial evaluation of its analytical tools.

Seven of the requests were refused, with the DWP citing Section 31 in these cases (of these, three requesters ask for an internal review). These requests pertain to general inquiries on AI models used to detect fraud, including their costs, effectiveness and the developer. One requester asked for information on the data and algorithm used for the General Matching Service. Two requesters asked for specific documents: equality impact assessments, DPIAs and evaluations of IRIS.

Finally, DWP did not hold information on two requests asking how DWP used AI to monitor social media, which DWP stated they are not doing.

FOIA Process Insights

In this phase, we find that requests for specific documents were more effective than general questions about data analytics for fraud detection. Requests for more general information are only partially answered or refused based on Section 31 and Section 24, while four requests for specific documents related to Universal Credit, the A&I Hub and third-party data, made between July 2019 and August 2021, yielded lightly redacted document releases.

Success was limited though. Starting in 2020, the DWP began leveraging Section 31 frequently to refuse document releases, particularly those relevant to IRIS-related queries. Evidence for this is based on three denied requests for IRIS-related documents made in Autumn of 2020; the DWP did not release training manuals, reports, equality impact assessments, nor DPIAs relevant to the program.

Ultimately, it required a complaint to the ICO and a subsequent investigation for the DWP to release heavily redacted versions of IRIS-related DPIAs in response to a February 2022 request (R18). Among

these is DPIA 1300 Part 2 v2.2, which describes the 'Common Risk Engine' that works to recuperate fraudulent advances claims, and the 'Advances Model,' discussed next. Because of redactions, the link between these two systems remains unclear.²⁸ This request spotlights that for some cases, the DWP will first refuse requests citing Section 31, then release heavily redacted documents after an internal review request or ICO complaint.

The Advances Model Phase, 2022-2023

Context

The second phase begins with the publication of DWP's *Annual Report and Accounts 2021-22*, which – for the first time – mentions DWP's use of risk modelling,²⁹ carried out by the department's Advances Model. The model 'analyses information from historical fraud cases to predict which cases are likely to be fraudulent in the future'.³⁰

The Advances Model is an example of profiling³¹ – it provides a risk score to applications for a Universal Credit advance; cases deemed at risk of fraud or error are referred to a case worker to determine if more checks need to be applied. Many applicants ask for this advance or loan to pay bills while they wait for their first Universal Credit payment, since initial payments can take up to six weeks after a successful application. Claimants then pay the advance back to the DWP, with increments taken out of each Universal Credit payment until paid back in full.³²

Prior to Covid-19, advances required face-to-face caseworker intervention. During Covid-19, when cases were no longer determined in-person, the DWP suspected high numbers of advances claims to be fraudulent.³³

²⁸ A tender for the Common Risk Engine shows that it was developed by a third-party, Kainos Software, for £3.3 million: <https://www.contractsfinder.service.gov.uk/Notice/be9bdeb6-d6de-4db8-8cf4-3320b9bfb2c1> and details on tender here: <https://bidstats.uk/tenders/2021/W50/764936495>

²⁹ Also called the Fraud Risk Model in the *2022-2023 Annual Report and Accounts*: <https://www.gov.uk/government/publications/dwp-annual-report-and-accounts-2022-to-2023/dwp-annual-report-and-accounts-2022-to-2023>

³⁰ <https://assets.publishing.service.gov.uk/media/62d52f55e90e071e814807fd/annual-report-accounts-2021-22-web-ready.pdf>

³¹ Profiling is legally covered under DWPs Personal Information Charter: <https://www.gov.uk/government/organisations/department-for-work-pensions/about/personal-information-charter#dwp-uses-of-profiling>

³² <https://www.gov.uk/guidance/universal-credit-advances>

³³ In 2019 DWP saw that advice on how to commit advances fraud was being shared on social media pages. <https://www.nao.org.uk/wp-content/uploads/2020/03/Universal-Credit-advances-fraud.pdf>

In November 2022, the House of Commons published a report on fraud and error in the benefits system. The report mentions that the DWP ‘has taken steps to evaluate the potential impact of data analytics and machine learning on groups with protected characteristics, but the results are inconclusive and it has not made them public.’ The report warns that ‘this lack of transparency over its use of data analytics risks eroding public trust in the benefit system.’³⁴

FOI Requests

There were 12 requests during this phase, with the first submitted in July 2022 and the last in February 2023.

Most requests during this time focus on the new Advances Model: how much it costs, any cost savings from its accurate fraud detection, the developer of the model and whether this model development was internal or involved third-party contracts. Requesters want to know how the model works and technical specifications: whether all claims or only a percentage are subject to the model, what data it uses, what type of model it is, how often the model is updated, source its code, and the distributions of its features and labels.

Many questions also relate to oversight. Requesters want access to the DWP’s pre-trial Fairness Analysis and to know the proportion of flagged claims in relation to protected characteristics. They want to know about its accuracy and error rate. People also asked for Equality Impact Assessments, DPIAs and other evaluations. A set of questions relates to human checks, namely the involvement of caseworkers and the Risk Review Team in the decision-making process.

One question relates to social media monitoring, asking about guidance in this area, DPIAs, information on the software used, the legal basis and auditing mechanisms in place.

DWP Responses

There are no successful requests in this phase. Nine are partially answered; in four of these, the requester made an internal review request.

From these nine, we find out some features of the Advances Model. Procedurally, the model identifies ‘advance applications that represent potentially increased risk of incorrectness or fraud;’ those with a high-risk score go to a member of staff to determine whether the claimant should be paid or asked to submit further information. Random samples that are not scored as high risk are also sent to caseworkers – without the caseworkers’ knowledge – for accuracy checks. Technically, we find out

³⁴ <https://committees.parliament.uk/publications/31513/documents/176815/default/>

only DWP staff developed and maintain the model, that the model has been trained on historical Universal Credit Advance claims and that it is a supervised classifier (a type of model).

Via one request made in 2023 (R32) we find out more about the internal structure of DWP's fraud review teams. The Risk Review Team uses data matching (but not the advances model) on claims already in payment. The Enhanced Checking Service is a separate team and part of DWP's Counter Fraud, Compliance & Debt strand; it takes referrals from 'data scans and staff at the new claim stage.' Both the Enhanced Checking Service and Risk Review Team are part of the wider 'Enhanced review team structure.'³⁵ We also find out that an Open Source team in the DWP investigates social media data; for information on this work, the requester is referred to the DWP's Fraud Investigations Staff Guide, last updated in 2019 (but removed from gov.uk in 2024³⁶).

Several requests, including those that are partially successful, ask for DPIAs and Equality Impact Assessments; these are refused except for two requests in which the requesters made complaints to the ICO (R24 and R30); these yielded two documents. The first, DPIA 1300 Part 2 v2.2, was released for the second time, as it assesses the advances model.³⁷ The second is DPIA 917 'RTG Programme – D&A Cross Cutting Services'. DPIA 917 mentions the Common Risk Engine and two other data-intensive services: the 'Uplifted analytics service', a 'cloud-based scalable analytical platform' that 'pool(s) information from customers,' and a 'Customer Event History Workstream' to 'derive data driven insights and predictive models to improve operational productivity, improve the customer experience, decrease fraud and error and enable a proactive risk-based analysis to increase efficiency and prioritisation of workload within DWP.'

The DWP rejected three of the requests during this phase; these asked for the Advances Model's Fairness Analysis, Equality Impact Assessment and DPIA and for general details about the model. Rejections cite Section 31 and Section 43, the commercial interests exemption.

FOIA Process Insights

In this phase, the DWP is initially unwilling to release any requested documents about the Advances Model. Regarding the two requests for which the DWP releases documents, DPIAs 1300 and 917, this happens only after an ICO complaint, and both documents are heavily redacted.

³⁵ Also mentioned on p. 100 of the 2022-23 ARA:

<https://assets.publishing.service.gov.uk/media/64a576d47a4c230013bba1e7/annual-report-accounts-2022-23-web-ready.pdf>

³⁶ <https://www.gov.uk/government/publications/fraud-investigations-staff-guide>. Privacy International carried out an investigation of the guide in 2021: <https://www.privacyinternational.org/long-read/4395/shedding-light-dwp-part-1-we-read-uk-welfare-agencys-995-page-guide-conducting>

³⁷ DWP refers to an earlier ICO decision regarding its right to redact the document:

<https://ico.org.uk/media2/migrated/decision-notice/4024689/ic-176118-m6q9.pdf>

As a result, several requesters in this phase experience disparate outcomes for the same request. Two requesters asking for DPIAs of the Advances Model are refused, while two others are ultimately supplied these documents after challenging decisions to gain this information.

Future Modelling Phase

Context

We mark the third phase with the release of the NAO's *Report on Accounts, Department for Work & Pensions* in July 2023, which mentions that the DWP is trialling four further uses of risk modelling. The DWP's *Annual Reports and Accounts 2022-2023* also introduces its Targeted Case Reviews (TCRs), 'where its agents look at a sample of claims in payment;' IRIS refers those 'identified as having potential incorrectness' for additional review by engaging with the claimant.³⁸

During this time, the DWP's data analytics techniques continue to be subjected to internal governance and safeguards. The same 2023 NAO report discloses that the DWP has subjected the Advances Model to a pre-launch fairness analysis to test for disproportionate impacts on people with protected characteristics.³⁹ The fairness analysis found 'some evidence of bias toward older claimants in some of the models.'⁴⁰ In an October 2023, Parliamentary question by MP Matthew Offord about 'the potential impact' of AI in the welfare system, MP Mims Davies responds that the 'DWP always ensures appropriate safeguards are in place for the proportionate, ethical, and legal use of data.'⁴¹

The Advances Model continues to come under public scrutiny at this time. Out of concern that groups with protected characteristics are being disparately impacted by the model, a House of Commons report published in December 2023 recommends that the DWP 'consider explicitly the impact of data analytics and machine learning on legitimate claims being delayed or reduced, the number of people affected, and whether this is affecting specific groups of people.'⁴² An article in the *The Guardian*,

³⁸https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1167946/annual-report-accounts-2022-23-web-ready.pdf. See also <https://www.gov.uk/government/publications/targeted-case-review-management-information/targeted-case-review-management-information#methodology-information>.

³⁹ <https://www.nao.org.uk/wp-content/uploads/2023/07/dwp-report-on-accounts-2022-23.pdf>

⁴⁰ The report says 'this bias is to be 'expected because fraudsters are more likely to say they are aged over 25 to claim a higher value benefit payment.' Ibid. p. 48.

⁴¹ <https://questions-statements.parliament.uk/written-questions/detail/2023-10-18/203281>

⁴² <https://committees.parliament.uk/publications/42434/documents/210942/default/>

published in December 2024, reported that the DWP found bias in the model based on age, disability, marital status and nationality.⁴³

The following year, in 2025, the House of Commons chided the DWP for its ongoing lack of disclosures on the Advances Model. Its report, *DWP Customer Service and Accounts 2023-24*, reveals that the DWP carried out another, post-hoc fairness impact assessment in 2024 that it has not yet released to the House.⁴⁴

Finally, a September 2024 press release reveals that the DWP is interested in third-party data sharing with banks. The press release announces 'new laws to be introduced to crack down on fraud' in a Fraud, Error and Debt Bill that would give the DWP powers to 'require banks and financial institutions to share data that may show indications of potential benefit overpayments' (called the 'Eligibility Verification Measure').⁴⁵

FOI Requests

There were 18 total requests during this time.

Requesters ask about the four new machine learning model pilots, requesting manuals and documents related to how each works and how they are monitored and evaluated. People ask for the release of any fairness analyses or information on how results impact people with protected characteristics, and information about internal safeguards ensuring the models operate without bias.

Requesters also want the DWP to release information on post-hoc analyses and monitoring of the Advances Model, asking for the latest fairness analyses, quality and accuracy analyses, Equality Impact Assessments and DPIAs. Questions continue about how the model works – such as whether it flags people before or after receiving the advances payment and whether it delays payment for those flagged – and how it relates to case worker decision-making.

Requesters ask for specific internal DWP procedural documents: the DWP data strategy, the data dictionary, data catalogue, open data strategy, data ethics framework, the guide to automated decision-making, and a Data Protection by Design document.

⁴³ The Guardian reporting is based on the FOI request 'AI Strategy Information' included in our dataset. <https://www.theguardian.com/society/2024/dec/06/revealed-bias-found-in-ai-system-used-to-detect-uk-benefits>

⁴⁴ <https://committees.parliament.uk/publications/46441/documents/235266/default/>

⁴⁵ <https://www.gov.uk/government/news/new-laws-to-be-introduced-to-crack-down-on-fraud>. Also a public third-party data impact assessment refers to two Proof of Concepts regarding access to banking data for the purpose of identifying fraud https://assets.publishing.service.gov.uk/media/6564bab01524e6000da10168/DWP_third_party_data_impact_assessment_november_2023.pdf

Some requests relate to DWP's proof of concept for data sharing with banks. People request oversight documents, including Equality Impact Assessments and DPIAs; one asks for a count of how many cases the proof of concept generates for people with protected characteristics and whether the data shared is used to automate data analytics.

DWP Responses

Four requests from this period are successful. The DWP responded to one initially by asking for a refined, more specific request; the DWP then answered that the Advances Model went live in May 2022 and that it does not flag claims for fraud investigation but rather 'as having the potential for a high risk of incorrectness,' after which a case worker either progresses the case for payment or requests more information about the claim. Another successful request asks for the DWP's data dictionary, open data strategy and other governance mechanisms; the DWP released an unredacted data glossary and 'Guidance on how to use the Data Ethics Framework.' Two other successful requests saw the release of the DWP's 'Guide to Automated Decision-making' and 'Data Protection by Design' documents.

Among the six partially answered, one requester asked for the DWP's AI strategy, AI inventory, AI risk framework, Data Strategy, the Fairness Impact Assessment and details about fairness metrics applied to the Advances Model (R49). The DWP cites Section 22 – an exemption for information intended for future publication – for these documents, but it provides a somewhat redacted February 2024 Advances Model Fairness Analysis. This is a major release, as it marks the first post-hoc analysis of a model's performance. As mentioned above, *The Guardian* covered news of this release in December 2024 as it reveals that the analysis finds bias against protected characteristics.⁴⁶

Two requests yielded documents after the requesters asked for internal review. Initially citing Section 31, the DWP ultimately released an unredacted version of its Data Strategy and a lightly redacted 'Equality Analysis Third Party Data Gathering Measure,' a heavily redacted DPIA related to the DWP's Banking Initiative, a proof of concept using banking information to detect capital and abroad fraud and error.

In another request (R39), the DWP released documents after an ICO complaint. Citing the October 2023 question to the DWP about AI and welfare,⁴⁷ the requester asked for documents related to safeguards where AI is used for fraud detection; this request yielded five DPIAs: 1300, 1442, 1978, 2638, 1993. The DWP initially cited exemption Section 21, claiming that the information was already in the public domain. In this tranche, the DWP again released DPIA 1300, but as a much less redacted

⁴⁶ <https://www.theguardian.com/society/2024/dec/06/revealed-bias-found-in-ai-system-used-to-detect-uk-benefits>

⁴⁷ <https://questions-statements.parliament.uk/written-questions/detail/2023-10-18/203281>

version. The other four DPIAs relate to DWP's new machine learning models: the Living Together Model, the Self-employed Earnings model, the Housing Model, and a model to detect undeclared capital.

One request focuses on Targeted Case Reviews and asks whether automation is involved in this process, mentioned in the DWP's '22-'23 *Annual Reports and Accounts*, along with any relevant Equality Impact Assessments or DPIAs. The DWP replies that Targeted Case Reviews use automation but not a model or machine learning. Upon a second, narrowed request for DPIAs, the DWP releases somewhat redacted DPIAs 2097 and 2319, both assessing Targeted Case Review.

In two requests, the DWP did not hold the information sought. One person asked how long a case worker must review a Universal Credit claim flagged by the Common Risk Engine; the DWP replies that there is no prescribed timescale. In another request from December 2023, the requester asked for DPIAs and Equality Impact Assessments for the four new models, but the DWP claims the exercise has 'not commenced.'

The DWP refuses six requests in this phase. In five instances, the department cites Section 31; these requests ask for more details about the four new machine learning models, about the error rate of the Advances model, and for the fairness analysis for the Advances Model mentioned by NAO. For a sixth request, the DWP cites Section 14 – for when a request is vexatious – to reject a requester asking for the department's Fraud Investigations Staff Guide, as it is now held on an 'internal intranet site' and is no longer a single document. In one request, the DWP asked the requester to refine the request, but there was no record that this had happened at the time of data collection.

FOIA Process Insights

As with the phase before, this period is marked by the DWP making divergent responses to similar requests. Most notably, the DWP rejects requests for the Advances Model Fairness Analysis or Equality Impact Assessment in two answers sent after the Advances Model Fairness Analysis was created in February 2024, though it releases this same document to another requester in August 2024.

Also, in keeping with former practices, for four requests the DWP only releases information after an internal review is requested or an ICO complaint is made. In the cases of DPIAs, these remain heavily redacted.

Conclusions

FOI requests offer a way to counter the persistent lack of public scrutiny over opaque data-intensive and automated technologies in the public sector.⁴⁸ By tracing FOI requests across time, we get a more coherent chronology of internal teams, technologies and programmes in use and in development within the DWP. We gain insight into how charities and journalists deploy FOIA as a democratic tool – how they react to the publication of government reports, use FOIA procedures to obtain more details, and release these findings to the public in articles and reports, in some cases, bringing more public oversight and more FOI requests by others.

Crucially, the FOI request dataset reveals inconsistencies across the DWP's responses to requests: those who accept the DWP's default refusals do not see the same information as those willing to pursue the matter with the ICO. Particularly, we note the DWP's inconsistent release of their DPIAs – key documents used in the design phase of data-intensive technologies to understand their potential impact on citizens. We also highlight that, within our dataset, the DWP only released one document auditing potential bias or disparate impact – the 2024 Advances Model Fairness Analysis, which assessed how the model performs on some protected characteristics – over the seven-year period under review; the department never released any fairness analyses related to their data matching services over this time.

We conclude this report with a call for the DWP to be more transparent and consistent in its releases of information about its data-intensive fraud detection practices. **Rather than disclose information in a non-standardised fashion, the DWP should make already released and other relevant documents available to the public through the UK government's Algorithmic Transparency Recording Standard Hub, a searchable database of algorithmic tools used by public agencies.**⁴⁹ Such reporting would make the need to hunt down documents in this 'good-faith mosaic' approach unnecessary.

⁴⁸ Scott, R. M.-G., & Edwards, L. (2025). The Inscrutable Code? The Deficient Scrutiny Problem of Automated Government. *Technology and Regulation*, 2025, 37–59. <https://doi.org/10.71265/sxb9dj82>

⁴⁹ <https://www.gov.uk/algorithmic-transparency-records>

References

- Big Brother Watch. (2021, July 20). Poverty panopticon. The hidden algorithms shaping Britain's welfare state. <https://bigbrotherwatch.org.uk/wp-content/uploads/2021/07/Poverty-Panopticon.pdf>
- Booth, R. (2023, September 3). UK warned over lack of transparency on use of AI to vet welfare claims. *The Guardian*. <https://www.theguardian.com/politics/2023/sep/03/uk-warned-over-lack-transparency-use-ai-vet-welfare-claims>
- Booth, R., & editor, R. B. U. technology. (2024, December 6). Revealed: Bias found in AI system used to detect UK benefits fraud. *The Guardian*. <https://www.theguardian.com/society/2024/dec/06/revealed-bias-found-in-ai-system-used-to-detect-uk-benefits>
- Burgess, M. (n.d.). This Algorithm Could Ruin Your Life. *Wired*. Retrieved 21 June 2025, from <https://www.wired.com/story/welfare-algorithms-discrimination/>
- Committee of Public Accounts. (2025). DWP Customer Service and Accounts 2023-24 (Sixth Report of Session 2024–25 HC 354). <https://committees.parliament.uk/publications/46441/documents/235266/default/>
- Dencik, L. (2022). The Datafied Welfare State: A Perspective from the UK. In A. Hepp, J. Jarke, & L. Kramp (Eds.), *New Perspectives in Critical Data Studies: The Ambivalences of Data Power* (pp. 145–165). Springer International Publishing. https://doi.org/10.1007/978-3-030-96180-0_7
- Department for Work & Pensions. (2018). 2018-2022 Fraud, Error and Debt Strategy Tackling fraud, error and debt in the benefit system. <https://pdacounterfraud.co.uk/wp-content/uploads/2023/07/DWP-Fraud-Error-and-Debt-Strategy-2018-2022.pdf>
- Department for Work & Pensions. (2018). Department for Work and Pensions Annual Report and Accounts 2017-2018. <https://assets.publishing.service.gov.uk/media/5b364091e5274a0bc32fabf7/dwp-annual-report-and-accounts-2017-2018.pdf>
- Department for Work & Pensions. (2020). Department for Work and Pensions Annual Report and Accounts 2019-2020 (HC 401).

<https://assets.publishing.service.gov.uk/media/5ef9f3b2d3bf7f769392421e/dwp-annual-report-and-accounts-2019-2020.pdf>

Department for Work & Pensions. (2022). Department for Work and Pensions Annual Report and Accounts 2021-2022. Department for Work and Pensions. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1091361/annual-report-accounts-2021-22-web-ready.pdf

Department for Work & Pensions. (2024, September 24). New laws to be introduced to crack down on fraud. GOV.UK. <https://www.gov.uk/government/news/new-laws-to-be-introduced-to-crack-down-on-fraud>

Department of Work & Pensions. (n.d.). Personal information charter. GOV.UK. Retrieved 27 June 2025, from <https://www.gov.uk/government/organisations/department-for-work-pensions/about/personal-information-charter>

Department of Work & Pensions. (2023). Department for Work and Pensions annual report and accounts 2022-2023. <https://www.gov.uk/government/publications/dwp-annual-report-and-accounts-2022-to-2023/dwp-annual-report-and-accounts-2022-to-2023>

Department of Work & Pensions. (2025a). Quality statement: DWP benefits statistical summary. <https://www.gov.uk/government/publications/dwp-statistical-summary-policies-and-statements/quality-statement-dwp-benefits-statistical-summary>

Department of Work & Pensions. (2025b, May 22). Targeted Case Review Management Information. GOV.UK. <https://www.gov.uk/government/publications/targeted-case-review-management-information/targeted-case-review-management-information>

Elyounes, D. (2021). "Computer Says No!": The Impact of Automation on the Discretionary Power of Public Officers. *Vanderbilt Journal of Entertainment & Technology Law*, 23(3), 451.

Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press.

Hamilton, M. (2017, December 30). Robots will scour Facebook for criminal activity in a £2bn benefit fraud blitz. *The Sun*. <https://www.thesun.co.uk/news/5238537/robots-will-scour-facebook-for-criminal-activity-in-a-2billion-benefit-fraud-blitz/>

HM Revenue and Customs, & Department for Work & Pensions. (2021). Tackling fraud and error in the benefit and tax credits systems. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/214334/tackling-fraud-and-error.pdf

House of Commons Committee of Public Accounts. (2022). The Department for Work and Pensions' Accounts 2021–22 – Fraud and error in the benefits system (Twenty-Sixth Report of Session 2022–23). <https://committees.parliament.uk/publications/31513/documents/176815/default/>

House of Commons Committee of Public Accounts. (2023). The Department for Work & Pensions Annual Report and Accounts 2022–23 (Fourth Report of Session 2023–24). <https://committees.parliament.uk/publications/42434/documents/210942/default/>

Marr, B. (n.d.). How The UK Government Uses Artificial Intelligence To Identify Welfare And State Benefits Fraud. Forbes. Retrieved 27 June 2025, from <https://www.forbes.com/sites/bernardmarr/2018/10/29/how-the-uk-government-uses-artificial-intelligence-to-identify-welfare-and-state-benefits-fraud/>

National Audit Office. (2018). Rolling out Universal Credit (HC 1123, 15 June Session 2017-2019). <https://www.nao.org.uk/reports/rolling-out-universal-credit/>

National Audit Office. (2020). Universal Credit advances fraud (HC 105 SESSION 2019–2021). <https://www.nao.org.uk/wp-content/uploads/2020/03/Universal-Credit-advances-fraud.pdf>

National Audit Office. (2023). Report on Accounts, Department for Work & Pensions. <https://www.nao.org.uk/wp-content/uploads/2023/07/dwp-report-on-accounts-2022-23.pdf>

National Audit Office. (2024). Report on Accounts, Department for Work & Pensions. National Audit Office. <https://www.nao.org.uk/wp-content/uploads/2024/07/dwp-report-on-accounts.pdf>

Pyper, D. (2020). The Public Sector Equality Duty and Equality Impact Assessments. <https://commonslibrary.parliament.uk/research-briefings/sn06591/>

Scott, R. M.-G., & Edwards, L. (2025). The Inscrutable Code? The Deficient Scrutiny Problem of Automated Government. *Technology and Regulation*, 2025, 37–59. <https://doi.org/10.71265/sxb9dj82>

UK Parliament. (n.d.). Written questions and answers—Written questions, answers and statements—UK Parliament. Retrieved 27 June 2025, from <https://questions-statements.parliament.uk/written-questions/detail/2023-10-18/203281>

UK Parliament. (2021a, April 23). Written questions and answers—UK Parliament—Social Security Benefits: Fraud, Question for Department for Work and Pensions. <https://questions-statements.parliament.uk/written-questions/detail/2021-04-23/186257/>

UK Parliament. (2021b, November 19). Written questions and answers—UK Parliament—Social Security Benefits: Disqualification, Question for Department for Work and Pensions. <https://questions-statements.parliament.uk/written-questions/detail/2021-11-19/78475/>

UK Parliament. (2022, January 26). DWP Risk Review Team—Hansard—UK Parliament. <https://hansard.parliament.uk/Commons/2022-01-26/debates/333BCD75-7B81-464A-BA13-91D711B1A4EF/DWPRiskReviewTeam>

Van Bekkum, M., & Borgesius, F. Z. (2021). Digital welfare fraud detection and the Dutch SyRI judgment. *European Journal of Social Security*, 23(4), 323–340. <https://doi.org/10.1177/13882627211031257>

Table I: FOI Requests Timeline

Case Number	Title	Date of First Submission	Date of Last Submission	Link
1	FRAUD-busting Robots/ Artificial Intelligence	9-Apr-18	18-Apr-19	https://www.whatdotheyknow.com/request/fraud_busting_robotsartificial_i
2	AI Trial to Combat Benefit Fraud	18-Nov-18	8-Mar-19	https://www.whatdotheyknow.com/request/ai_trial_to_combat_benefit_fraud#incoming-1291183
3	False Positives in AI Trial to Prevent Fraud	19-Nov-18	3-Jan-19	https://www.whatdotheyknow.com/request/false_positives_in_ai_trial_to_p#incoming-1287278
4	Data Sources and Processes	24-Feb-19	18-Jun-19	https://www.whatdotheyknow.com/request/data_source_s_and_processes
5	Universal Credit-related PIAs, DPIAs, DSAs and MoUs	16-Jul-19	15-Aug-19	https://www.whatdotheyknow.com/request/universal_credit_related_pias_dp#incoming-1419505
6	Analysis and Intelligence Hub for Universal Credit	23-Jul-19	18-Oct-19	https://www.whatdotheyknow.com/request/analysis_and_intelligence_hub_fo#incoming-1419975
7	Request for Corporate Documents	27-Nov-19	30-Jan-20	https://www.whatdotheyknow.com/request/request_for_corporate_documents#incoming-1513472
8	Social Media Monitoring	23-Jan-20	18-Feb-20	https://www.whatdotheyknow.com/request/social_media_monitoring_254#incoming-1529097
9	DWP's Welfare Fraud and Error Detection Programs	23-Jan-20	19-Feb-20	https://www.whatdotheyknow.com/request/dwps_welfare_fraud_and_error_det#incoming-1529750
10	Automated Risk Analysis and Intelligence System	28-Aug-20	21-Oct-20	https://www.whatdotheyknow.com/request/automated_risk_analysis_and_inte#outgoing-1065691
11	DWP Fraud and Error Detection Tools	11-Dec-20	4-Jan-21	https://www.whatdotheyknow.com/request/dwp_fraud_and_error_detection_to#incoming-1698461
12	Digital Economy Act project with the Student Loan Company	8-Aug-21	7-Sep-21	https://www.whatdotheyknow.com/request/digital_economy_act_project_with#incoming-1870565
13	Impact assessments for the General matching service	28-Oct-21	25-Nov-21	https://www.whatdotheyknow.com/request/impact_assessments_for_the_gener
14	General Matching Service and its relationship to the National Fraud Initiative	28-Oct-21	11-Nov-21	https://www.whatdotheyknow.com/request/general_matching_service_and_its
15	The DWP's General Matching System and its relation to fraud	28-Oct-21	25-Nov-21	https://www.whatdotheyknow.com/request/the_dwps_general_matching_system#incoming-1922889
16	Design of and Data Use in the General Matching Service	28-Oct-21	25-Nov-21	https://www.whatdotheyknow.com/request/design_of_and_data_use_in_the_ge#incoming-1923016
17	AI in Use with DWP	29-Nov-21	23-Dec-21	https://www.whatdotheyknow.com/request/ai_in_use_with_dwp#incoming-1942218
18	IRIS Data Protection Impact Assessments	6-Feb-22	21-Feb-23	https://www.whatdotheyknow.com/request/iris_data_protection_impact_asse#incoming-2241628
19	Monitoring and Analysis Service for Fraud and Error	10-Feb-22	28-Feb-22	https://www.whatdotheyknow.com/request/monitoring_and_analysis_service#incoming-1984487
20	New counter-fraud plan - arrests/warrants/searches/evidence seizure/banks	20-May-22	9-Jun-2022	https://www.whatdotheyknow.com/request/new_counter_fraud_plan_arrestswa#incoming-2056789
21	Risk Review Team's Use of Automation	25-May-22	30-Aug-22	https://www.whatdotheyknow.com/request/risk_review_teams_use_of_automat#incoming-2111441
22	Fairness Analysis	8-Jul-22	6-Sep-22	https://www.whatdotheyknow.com/request/fairness_analysis#incoming-2097896
23	Universal Credit Algorithm	12-Jul-22	2-Aug-22	https://www.whatdotheyknow.com/request/universal_credit_algorithm#incoming-2093115
24	Universal Credit Fraud Algorithm	13-Jul-22	15-Sep-23	https://www.whatdotheyknow.com/request/universal_credit_fraud_algorithm#incoming-2327475
25	Universal Credit Advances Fraud Risk Model-Machine Learning Algorithm	15-Jul-22	29-Sep-22	https://www.whatdotheyknow.com/request/universal_credit_advances_fraud#incoming-2135526

26	Freedom of Information Request in Relation to Social Media Intelligence	26-Jul-22	24-Aug-22	https://www.whatdotheyknow.com/request/freedom_of_information_request_i_80#incoming-2107971
27	Fraud Risk Model - Universal Credit Advance Claims	5-Aug-22	21-Nov-22	https://www.whatdotheyknow.com/request/fraud_risk_model_universal_credi#incoming-2113867
28	Risk Model DPIA	5-Oct-22	2-Nov-22	https://www.whatdotheyknow.com/request/risk_model_dpia#incoming-2158335
29	Universal Credit Advance Claims – fraud risk model	24-Oct-22	6-Feb-23	https://www.whatdotheyknow.com/request/universal_credit_advance_claims#incoming-2173661
30	Iris Common Risk Engine	1-Feb-23	15-Sep-23	https://www.whatdotheyknow.com/request/iris_common_risk_engine#incoming-2420044
31	Universal Credit Fraud Risk Model and Machine Learning Algorithm	17-Feb-23	9-Mar-23	https://www.whatdotheyknow.com/request/universal_credit_fraud_risk_mode#incoming-2256450
32	Risk Review Team	17-Feb-23	30-May-23	https://www.whatdotheyknow.com/request/risk_review_team#incoming-2262635
33	Risk Model to Detect Fraud	24-Feb-23	10-Mar-23	https://www.whatdotheyknow.com/request/risk_model_to_detect_fraud#incoming-2258179
34	Artificial Intelligence in Preventing UC Fraud	12-Jul-23	26-Jul-23	https://www.whatdotheyknow.com/request/artificial_intelligence_in_preve
35	Pilot of Four New Fraud Prevention Models	1-Aug-23	31-Aug-23	https://www.whatdotheyknow.com/request/pilot_of_four_new_fraud_preventi#incoming-2405325
36	Targeted Case Reviews Process	1-Aug-23	24-Oct-23	https://www.whatdotheyknow.com/request/targeted_case_reviews_process#incoming-2452498
37	Fairness Analysis of Data Analytics Products	2-Aug-23	24-Aug-23	https://www.whatdotheyknow.com/request/fairness_analysis_of_data_analyt#incoming-2401106
38	DWP Data Strategy	7-Aug-23	19-Oct-23	https://www.whatdotheyknow.com/request/dwp_data_strategy#incoming-2449199
39	AI Internal Monitoring Protocols	25-Oct-23	15-Mar-24	https://www.whatdotheyknow.com/request/ai_internal_monitoring_protocols#incoming-2591451
40	Caseworker involvement in universal credit fraud investigations	31-Oct-23	16-Nov-23	https://www.whatdotheyknow.com/request/caseworker_involvement_in_univer/response/2471021/attach/html/3/Response%20FOI%202023%2083524.pdf.html
41	Documents Relating to DWP Use of Data	5-Dec-23	5-Feb-24	https://www.whatdotheyknow.com/request/documents_relating_to_dwp_use_of#incoming-2542992
42	Banking Data - Proof of Concepts	19-Dec-23	16-Feb-24	https://www.whatdotheyknow.com/request/banking_data_proof_of_concepts
43	DWP Third Party Data Gathering	19-Dec-23	15-Jan-24	https://www.whatdotheyknow.com/request/dwp_third_party_data_gathering#incoming-2520190
44	Guide to Automated Decision-making	3-Jan-24	18-Jan-24	https://www.whatdotheyknow.com/request/guide_to_automated_decision_maki
45	Data Protection by Design	18-Jan-24	20-Feb-24	https://www.whatdotheyknow.com/request/data_protection_by_design#incoming-2536761
46	Universal Credit Advances Claims - Fraud Risk Model	21-Feb-24	15-April-24	https://www.whatdotheyknow.com/request/universal_credit_advance_claims#incoming-2173661
47	Impact and Accuracy of Universal Credit Advances Claims Model	23-Feb-24	18-Mar-24	https://www.whatdotheyknow.com/request/impact_and_accuracy_of_universal#incoming-2593484
48	DWP Fraud Investigations Staff Guide	21-May-24	19-Dec-24	https://www.whatdotheyknow.com/request/dwp_fraud_investigations_staff_g#incoming-2866895
49	AI Strategy Information	2-Aug-2024	2-Sep-24	https://www.whatdotheyknow.com/request/ai_strategy_information#incoming-2748592
50	Equality Analysis for Fraud, Error and Debt Bill Eligibility Verification Measure and Data Protection and Digital Information Bill Third Party Data (TPD) Gathering Measure	18-Oct-24	4-Feb-25	https://www.whatdotheyknow.com/request/equality_analysis_for_fraud_erro#outgoing-1785775
51	What is the Known Error Rate of the DWP Algorithm?	7-Dec-24	14-Jan-25	https://www.whatdotheyknow.com/request/what_is_the_known_error_rate_of

Table II: DWP Response Table

	Total Requests	Successful	# Docs Released	Partially Successful	# Docs Released	Refused	Information Not Held
Phase 1	21	3	0	9	5	7	2
Phase 2	12	0	0	9	2	3	0
Phase 3	18	4	3	6	5	6	2

Table III: Released Documents

Title	Redaction status	FOI ID No.
Privacy Impact Assessment v4.3 'Living Together Data & Analytics' project	Lightly	5
DWP Data Sharing Agreement Draft - Land Registry Property Ownership Data Pilot	Lightly	5, 9
DWP Data Sharing Agreement Draft - Universal Credit - data matching to Student loans pilot	Lightly	5, 9
MOU - Universal Credit data match with Student Loan Company within the NFI	Lightly	5
MOU between DWP and HMRC	Lightly	5
MOU between the Home Office and DWP	Lightly	5
MOU between the Home Office, HMRC and DWP	Lightly	5
MOU between HMRC and DWP and Department for Communities	Lightly	5
MOU between DWP and HMRC in Respect of Data Sharing of Personal Information with Data and Analytics	Lightly	5
'DWP Data & Analytics Directorate' org chart	Unredacted	7
Data Protection Governance Board Terms of Reference	Unredacted	7
Data Protection Governance Board Meeting 18 July 2019	Unredacted	7
Data Protection Governance Board Meeting 17 October 2019	Unredacted	7
Data Protection Governance Board Meeting 17 September 2019	Unredacted	7
PIA - Land Registry Property Ownership Data Pilot	Lightly	9
DPIA - DVLA and DWP data sharing for UC verification	Lightly	9
DWP Data Sharing Agreement - Universal Credit verification and fraud detection pilot	Lightly	9
Digital Economy Act 2017 Debt and Fraud Information Sharing Business Case - Department for Work and Pensions (DWP) Student Loan Company (SLC) - Universal Credit/Student Loan Fraud Scoping	Lightly	12
DWP Data Sharing Agreement (DSA) between The Secretary of State for Work and Pensions and The Student Loans Company Limited	Lightly	12
DPIA 959 - Student Loan Company	Lightly	12

DPIA No number part 2 v3.1 provided	Heavily redacted	18
DPIA 575-2 part 2 v2.1	Heavily redacted	18
DPIA 978 part 2 v3.0	Heavily redacted	18
DPIA 1442 part 2 v3.1	Heavily redacted	18
DPIA 791 part 2 v2.1	Heavily redacted	18
DPIA 1300 part 2 v2.2	Heavily redacted	18, 24, 39
DPIA 1026 part 2 v2.2	Heavily redacted	18
DPIA 278 V8.0 part 2 v1.8	Heavily redacted	18
DPIA 917 'RTG Programme – D&A Cross Cutting Services'	Heavily redacted (two versions)	30
DPIA 2097 v3.2	Heavily redacted	36
DPIA 2319 v3.2	Heavily redacted	36
DWP Data strategy	Unredacted	38
DPIA 1442 'Living together model'	Heavily redacted	39
DPIA 1978 'self employed earnings model'	Heavily redacted	39
DPIA 2638 'housing model'	Heavily redacted	39
DPIA 1993 'undeclared capital'	Heavily redacted	39
Data glossary	Unredacted	41
Data Ethics Framework	Unredacted	41
Privacy Impact Assessment 497A v4 DWP Banking Initiative	Heavily redacted	42
DPIA 2134 'Aggregated Statistics to support Third Party Data Gathering Measure'	Heavily redacted	42
'Guide to Auto. Decision-making'	Unredacted	44
Data Protection by Design'	Unredacted	44
Advances Model Fairness Analysis Summary Report February 2024	Somewhat redacted	49
Equality Analysis Third Party Data Gathering Measure 29/08/2023	Somewhat redacted	50

Advances Model: a supervised machine learning model trained on data of past fraud cases to predict future cases that are likely to be fraudulent. The model gives a risk score to claims to assess fraud-and error risk of Universal Credit advances applications; the model has been in use since 2022.

Analysis and Intelligence (A&I) Hub: The A&I Hub deploys rules using DWP and third-party data sources to identify inconsistencies in benefit entitlement for Universal Credit claimants. The hub uses data-driven rules to determine if a Universal Credit transaction should continue or if it should be flagged for further human checks.

Common Risk Engine: used by IRIS to identify both individual and organised fraud and error. The Common Risk Engine entails machine learning and predictive modelling with a case management system to assess claims for risk.

Data Protection Impact Assessment (DPIA): This document is a legal requirement of GDPR and designed to help an organisation analyse, identify and minimise data protection risks. Government projects requiring the processing of personal data must fill out a DPIA form before project commencement and submit to the Information Commissioner's Office.

Equality Impact Assessment: Non-compulsory assessment completed by public authorities to assess whether a policy will have a disparate impact on persons with protected characteristics and to comply with equality duties.

Enhanced Checking Service: According to DWP's Annual Report & Accounts 2020-21, this is a 'team of trained investigators who review claims and contact claimants by telephone in order to obtain further information or evidence where there is suspected fraud.' According to R21, the enhanced checking service takes referrals at the new claim stage from both data scans and staff.

Enhanced Review Team: A DWP team dedicated to ongoing review of Universal Credit benefits to determine fraud and error; the Enhanced Review Team includes the Risk Review Team and the Enhanced Checking Service.

Advances Model Fairness Analysis: the DWP's Advances Model Fairness Analysis Summary Report from February 2024 appears to be an Equality Impact Assessment; it assesses the model's impact on protected characteristics of age, disability, marital status and nationality.

General Matching Service: A tool for identifying potential fraud or error in benefits cases and ensuring coherence and consistency in DWP data. Provides rules to find inconsistencies between a claimant's case and the data held by the DWP.

Internal Review: A complaints procedure that can be initiated if a requester is dissatisfied with the response content or handling of their Freedom of Information request.

Integrated Risk and Intelligence Service (IRIS): A service launched by DWP in 2022 to identify and prevent cases of fraud and error through data matching and data analytics. IRIS combines the Risk and Intelligence Service with DWP's other intelligence services, such as those detecting cybercrime.

Information Commissioner's Office FOI complaint: A person submitting a FOI request can challenge an agency's answer by requesting an assessment by the ICO to determine if an agency is following the Freedom of Information Act.

Open-Source Team: A DWP team that investigates social media data of claimants suspected of committing fraud.

Risk and Intelligence Service (RIS): Started in 2018/19, RIS uses data to prevent fraud and error and improve the efficiency of DWP's debt collection activity.

Risk Review Team: A team that investigates claims already in payment that have been flagged by the Integrated Risk and Intelligence Service (IRIS) as being a high fraud risk.

Targeted Case Review: A process established in 2022 to identify and address fraud and error across Universal Credit claims. Targeted Case Review entails over 5000 reviewing agents in the DWP who look at a sample of claims in payment that were automatically flagged for error; IRIS refers those for additional review by engaging with the claimant.