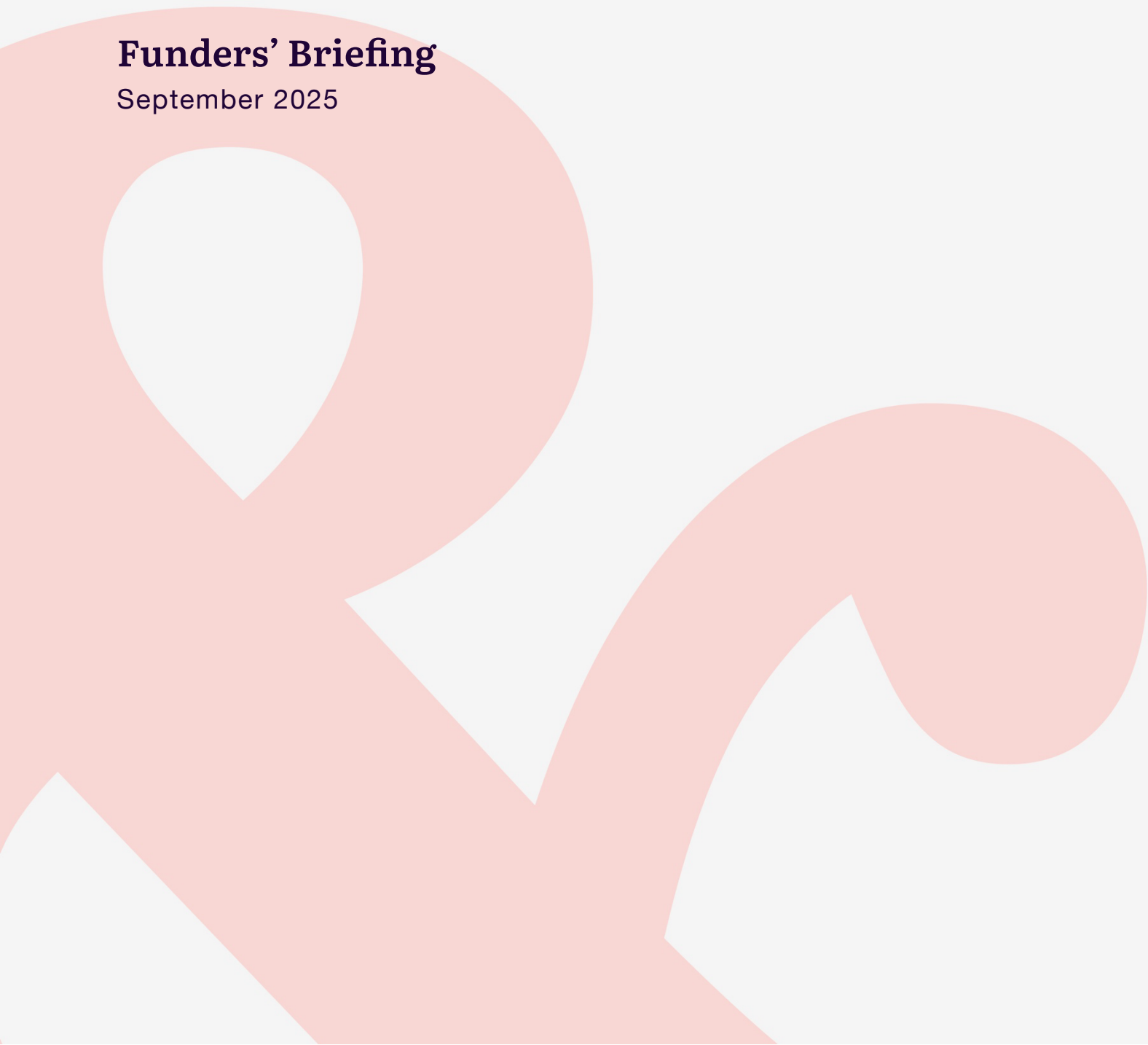



AI & Social Welfare

Funders' Briefing

September 2025



Executive Summary



Governments across Europe are using Artificial Intelligence (AI) to decide who gets access to social support. But technologies that promise greater efficiency in the distribution of public funds are in practice reinforcing discrimination and bias, harming some of the most vulnerable people in our societies, and leaving thousands of families destitute.

Cuts to benefits, intrusive home visits to people wrongly flagged as 'high risk' for welfare fraud, and discrimination are some of the tangible ways that AI-driven transformation is already affecting people's lives, cementing inequalities and entrenching injustice.

Several of the European AI & Society Fund's grantee community have been effectively investigating, litigating and campaigning against algorithmic injustice in welfare systems and providing guidance on the responsible deployment of AI. But they need longer-term support.

Philanthropy has an urgent role to play to challenge the negative impacts of AI-driven welfare systems and help direct AI use to better serve people and society. Funders that are committed to building just and fair societies can act by building the field of diverse public interest organisations fighting these harms.

Background

As governments across Europe digitise their public services, they are increasingly adopting AI and other automated decision-making tools into their social welfare systems. The European Union alone has committed 30 billion euros towards this transformation.

Social security payments are a large and growing part of the budget in most European countries, taking up almost 40% of EU member states' budgets in 2023. AI is presented as a way to cut this bill by making the system more efficient.

With media and public attention focused on addressing the perceived abuse of taxpayer funded systems – particularly by migrants and minorities – AI is being adopted to flag if someone is likely to commit social security fraud, with systems rolled out in France, Denmark and the Netherlands, among other countries. For example, the UK's Department of Work and Pensions (DWP) invested £70 million in advanced analytics between 2022–25, to tackle fraud and error. DWP projects these efforts will result in savings of £1.6bn by 2031. But often these algorithms are discriminatory, flagging people as 'high risk' for welfare fraud due to biased or incomplete data. Citizens become guilty until proven innocent because a system decided they *might* commit fraud.

Meanwhile AI is also deciding who is eligible for welfare support. In the name of efficiency and increasing demand on social security systems, AI is being used to assess people's complex welfare needs and life situations with binary algorithms and inbred bias. Hundreds of thousands of European citizens who were receiving social security support have been cut off, while others are denied their right to claim. Welfare workers are now quality assessors of opaque AI decisions, removed from direct contact with citizens. This dehumanises the social security system, degrading empathy and making it easier to discriminate without accountability. It is in this context that some civil society organisations have uncovered systemic AI-driven welfare scandals. This report summarises some of them and what philanthropists can do to help.

3 problems with AI & Welfare

Artificial Intelligence systems generate outputs based on the data that is inputted. In the case of welfare systems this creates several problems.

1) AI-powered welfare systems are often built on bad data, and data for marginalised communities is even lower quality than for other groups.

Investigations by Amnesty International found that the Serbian Social Card registry, funded by the World Bank, was built on out of date and incomplete government data. When it was rolled out, payments for some people were suddenly stopped, pushing already vulnerable people deeper into poverty. This is because much of the source data collected by the registry was incorrect, misrepresenting their income status or attributing income they had never had.

In Serbia, Mirjana was disqualified from receiving social assistance after a local human rights organisation helped to cover the cost of her daughter's funeral, who died unexpectedly in 2023. The organisation's donation of 20,000 Serbian dinars (around 170) into Mirjana's bank account was instantly flagged by the Social Card registry as income that disqualified her from social assistance. Practically overnight, Mirjana, who had been surviving on modest welfare benefits and living in a social housing complex, lost her assistance. Within two months of losing her daughter, Mirjana found herself in a prolonged and uncertain bureaucratic battle to reinstate her lost benefits.

Source: "World Bank-funded digital welfare system exacerbating poverty, especially for Roma and people with disabilities.", Amnesty International, December 2023.

2) Historic bias is replicated, and data doesn't capture the complexity of people's lives.

Fraud detection systems are set up to flag 'unusual' situations. Amnesty International's investigations in Denmark found that data-driven fraud control algorithms have been discriminating against low-income groups, racialised groups, migrants, refugees, ethnic minorities, people with disabilities, and older people. This is because 'unusual' living situations such as multi-occupancy, intergenerational households, and 'foreign affiliations' have flagged people as higher-risk of benefit fraud, based on data from previous practices. If

communities were discriminated against in the past, this is replicated and amplified in the systems built on that information

Source: "Coded Injustice: Surveillance and Discrimination in Denmark's automated welfare state", Amnesty International, November 2024.

"The way the Danish automated welfare system operates is eroding individual privacy and undermining human dignity. By deploying fraud control algorithms and traditional surveillance methods to identify social benefits fraud, the authorities are enabling and expanding digitised mass surveillance."

Hellen Mukiri-Smith
Amnesty International's researcher
on artificial intelligence and human rights

In the Netherlands, Lighthouse Reports found that a machine-learning algorithm generated a risk score to catch potential lawbreakers. It used limited, crude data about individuals who have been investigated for fraud in the past – including protected characteristics like gender, relationship status, age and whether someone was a parent or not – to assess someone's risk of welfare fraud. The complexity of their situation and life was not grasped or considered by AI. And it inherently discriminated against women, people with children, non-Dutch speakers, and migrants.

"Rotterdam's welfare-fraud algorithm is trained on millions of data points, but it is blind to what is actually happening in the real world. In reality, different people commit different kinds of violations – from honest mistakes when filling out a form to organised crime."

Source: "Inside the Suspicion Machine", a joint investigative report by Lighthouse Reports and WIRED magazine, 6 March 2023.

3) AI systems are deployed without transparency. People affected by their decisions are kept in the dark and have little chance of appealing.

For more than a decade and without any public consultation, France's AI-powered social welfare system CAF has deployed machine learning at a massive scale in a hunt for welfare fraud. Lighthouse Reports investigations in 2021 revealed that each year, nearly half of France's population is silently ranked by a secretive risk score with 13 million households flagged as high-risk for benefit fraud, due to them being people with disabilities, single mothers, and people with low income.

Hundreds of thousands of people had their benefits automatically cancelled, while others had fraud investigators come to their homes to count toothbrushes, question their neighbours and check bank accounts.

Source: "France's Digital Inquisition", Lighthouse Reports, December 2023.

Opportunities to challenge harmful AI-powered welfare systems

Although governments are embracing these technologies enthusiastically, many are still in their early stages of experimentation. Acting now to challenge harmful practices can prevent negative effects from getting baked into systems in the long-term. Instead, we can work towards welfare systems that work effectively and fairly to build strong and just societies for everyone.

Despite the opaque nature of these systems, investigations, litigation, community building and providing rights-respecting guidance have proved effective in driving change. Existing data protection and human rights legislation, as well as new laws like the European Union's AI Act, provide potentially effective routes to challenge harms.



Amnesty International – Uncovering systemic AI discrimination against non-Dutch nationals and denying childcare benefits

In the Netherlands, Amnesty International published a report about a childcare benefit scandal in 2021 which shook the country and contributed to the prime minister and other government officials being ousted. Amnesty Tech found that since 2013, the Dutch tax authorities had been using an algorithmic decision-making system to profile individuals applying for childcare benefits, to detect inaccurate and potentially fraudulent applications at an early stage. The system classified any non-Dutch national as 'higher risk', reinforcing the bias between race, ethnicity and crime. Tens of thousands of parents and caregivers were falsely accused of fraud and had their child benefits suspended. They were subjected to hostile investigations, characterised by harsh rules and policies, rigid interpretations of laws, and ruthless benefits recovery policies. This led to devastating financial problems for the families affected, ranging from debt and unemployment to forced evictions because people were unable to pay their rent or make payments on their mortgages. Others were left with mental health issues and stress on their personal relationships, leading to divorces and broken homes.

Amnesty's investigation made [national headlines](#), prompted affected parents to demand, and receive, compensation, and led to a parliamentary inquiry which ultimately ousted the prime minister and other senior government officials.

CIVIO - Winning a Supreme Court case for algorithmic transparency

Spanish journalists and campaigners Civio conducted a six-year long investigation into Spain's AI automated decision-making system BOSCO, which reviews applications from hundreds of thousands of vulnerable people – such as older people and those on lower-income – for help with energy bills. Civio exposed errors in the system which have led to certain eligible people being denied help.

Civio asked the Spanish government for more transparency around how the AI system works and access to its source code. In 2025, the Supreme Court ruled in Civio's favour, creating jurisprudence, and establishing that knowing the ins and outs of the programs and algorithms used by public administrations is a democratic right.

Source: ["Civio pulls back the curtain on public algorithms: Spain's Supreme Court orders the Government to release BOSCO's source code", Civio, September 2025.](#)

Algorithm Audit – Guiding governments to use algorithms responsibly

An investigation by Lighthouse Reports revealed how the benefit fraud detection algorithm in the Dutch city of Rotterdam had been disproportionately flagging claims by young single mothers with poor Dutch language skills as fraudulent. Their articles highlighted the devastating costs for the families who were wrongfully accused and had their benefits paused. In response, Algorithm Audit, a Dutch NGO that has provided ethical-AI auditing services to government ministries created a guide of strict conditions for the responsible use of these kinds of algorithms by public bodies. Developed through deliberation and testing to explore what a fair algorithm for benefit fraud detection looks like, the guide has been accepted by the Dutch ministry for digitalisation and provides a normative standard that can be used in all 342 municipalities in the Netherlands.

Source: [Risk Profiling for Social Welfare Reexamination, Algorithm Audit, 2023](#)

We understand the harms.
Now organisations need more capacity to investigate,
take action and guide governments to run better welfare systems.

What funders can do

Funders that are committed to building just and fair societies can contribute by supporting diverse public interest organisations that are already actively fighting these harms and by helping build the field for this work.

The European AI & Society Fund supports diverse public interest groups that are at the forefront of this new and fast-growing field. Submissions to our recent call for proposals for our €2 million AI Accountability grants demonstrate there is a huge appetite among civil society organisations to take on this work from diverse groups right across Europe – a total of €46 million was requested across 325 applications. We are now building a community of practice to deepen our understanding of how to challenge AI harms.

Support civil society to secure AI accountability in welfare systems

The European AI & Society Fund pools resources from 17 funding partners to support civil society organisations with:

Investigations	Understanding how AI is used in welfare systems requires socio-technical expertise to investigate the underlying code and algorithms to gather evidence of how the system affects individuals and communities. Research can take many years and requires persistence to gain transparency over opaque processes.	<u>Lighthouse Reports</u> , <u>Civio</u> , <u>Amnesty International</u>
Litigation	Evidence unearthed in investigations combined with careful legal analysis can challenge harmful deployments of AI systems.	<u>Amnesty International</u> , <u>Foxglove</u>
Community building	It's vital that the people most affected by the impacts of AI systems are in the driving seat of any movement for change.	<u>Weaving Liberation</u> , <u>Migrant Justice</u> <u>Community of Practice</u> <u>coalition</u> , <u>#ProtectNotSurveil</u> <u>coalition</u>
Providing guidance	Helping public authorities to use AI wisely and well could prevent the harmful impacts of AI in welfare systems in the future.	<u>Algorithm Audit</u>

What next?

AI may not be a funding priority for you. However, if health equity, human rights, social, migrant or disability justice matter for your work, then the intersection of AI and welfare should be a focus. Discriminatory AI systems are affecting the most vulnerable in society who often don't know how to challenge the injustice they are facing, and don't realise it is so widespread.

The European AI & Society Fund team can help you find the right strategy for your organisation's approach to funding around AI's impacts on society. You can support our work directly as one of our funding partners, or we can advise on relevant organisations within our network that could be a good fit.

Get in touch with our Senior Partnerships Manager Peggye to find out more:
peggye.totozafy@europeanaifund.org

About the European AI & Society Fund

The European AI & Society Fund mobilises a community of funders to fight for Artificial Intelligence that promotes fair, inclusive and sustainable societies. We support diverse public interest groups across Europe by galvanising funding to empower them with the capacity and skills to shape the AI agenda, free from corporate or government ties. Since 2020, we've awarded €10.5 million to 65+ organisations across 26 countries to shape AI to better serve people & society.

→ **Discover our Grantee Partners**

European
Artificial Intelligence
& Society Fund