14. Fundamental Rights and Algorithms Impact Assessment: Towards a More Inclusive and Accountable Digital Governance: Interview with Janneke Gerards

Viktorija Morozovaite

Introduction

Big data analyses and artificial intelligence systems are increasingly being used in public administration contexts around the world. While the digitalization of public decision-making processes offers opportunities for faster, more efficient, and consistent outcomes, it also creates risks related to equality, fairness, accountability, and other unintended consequences (Reisman et al. 2018). In the Netherlands, the impact of algorithmic decision-making systems on citizens and society at large has been widely discussed as part of the Dutch Digitalisation Strategy 2021 (Nederland Digitaal 2021; Van Til 2019). At the request of the Ministry of the Interior and Kingdom Relations, Prof. Janneke Gerards, together with Dr. Mirko Tobias Schäfer, Iris Muis, and Arthur Vankan of Utrecht University's Data School, developed the Fundamental Rights and Algorithms Impact Assessment (FRAIA)—an instrument to help identify and address the human rights risks posed by algorithms used by public organizations.

This section includes an interview with Janneke Gerards, professor of fundamental rights law at the Utrecht University, about the development of FRAIA and the lessons to be drawn for the European digital society.

Viktorija Morozovaite is **VM**, Janneke Gerards is **JG**.

Van Dijck, José, Karin van Es, Anne Helmond, and Fernando van der Vlist, eds. *Governing the Digital Society: Platforms, Artificial Intelligence, and Public Values*. Amsterdam: Amsterdam University Press, 2025
DOI 10.5117/9789048562718_CH14

VM: *Let's start by taking a step back to the year 2021. Could you explain the context in which FRAIA was developed? What concerns and risks related to fundamental rights in the Dutch and the broader European digital society inspired its development?*

JG: Algorithms have been around for quite some time, but around 2020, public authorities in the Netherlands started to realize that there can be many fundamental rights' risks involved. At the time, everyone was aware of data protection issues, but there were increasing examples of fundamental rights' risks coupled with an increasing use of algorithms by governmental organizations. For instance, the System Risk Indication (SyRI) algorithm (a risk-management algorithm, used by the Dutch government to predict the risk of detecting social security fraud) created a lot of debate because it had discriminatory aspects to it, and while everyone was thinking about data protection, non-discrimination concerns suddenly became relevant (Wieringa 2023).

At the time, the Dutch Ministry of the Interior and Kingdom Relations was developing a program on the various risks related to the use of algorithmic systems. As a first step, they needed some kind of inventory of the fundamental rights problems related to big data analyses and algorithms. They asked me to do that and together with research master student Max Vetzo and Professor Remco Nehmelman, we tried to identify all fundamental rights risks that could be relevant and showed the relevant legal framework.

As a follow-up on this, the ministry then asked me to develop a FRAIA-like instrument. The background to this was that the ministry saw public institutions jump on the train of adopting big data analyses, algorithmic systems, and AI technologies, without being fully aware of the risks. They felt there was a clear need for policy instruments to ensure that this was done in a streamlined and responsible manner.

Initially, the ministry asked us to design an assessment that purely focused on fundamental rights assessment, but it very quickly turned out that it would be much more useful to have a more holistic instrument. Such an instrument could include a variety of elements of ethical and responsible AI, as well as insights from political and governance sciences on the legitimacy and quality of public bodies' decision-making processes. To allow me to design a useful and workable model, I then joined forces with Utrecht University's Data School, which employs a number of great people working on how algorithmization and datafication transform democracy, and which has also developed the Data Ethics Decision Aid (DEDA). The

ministry agreed with this approach, and the collaboration resulted in the development of the design of such a sizeable and functional instrument.

VM: *In essence, FRAIA is an accountability mechanism, aimed to reduce the risks of carelessness, ineffectiveness, and infringement of the fundamental rights of citizens when governmental organizations implement algorithmic systems in their activities. How does FRAIA contribute to algorithmic accountability, and what safeguards are in place to ensure that it is not just a paper tiger?*

JG: Well, understandably, people can be wary of having to do lots of paperwork and some civil servants can be apprehensive of instruments that are changing the way they would normally work. Therefore, in designing FRAIA, we tried to logically follow the kind of processes that would normally happen in policy- and decision-making. Civil servants would have to go through the steps similar to those introduced by FRAIA anyway, they would have the discussions and make the decisions, but now they also write down the answers they give to all questions they encounter. We have done several practice tests before introducing FRAIA just to see how it would work in real-life settings, and so far, the results are quite positive.

In terms of accountability, I think FRAIA generates highly valuable results. By participating in the FRAIA process, in fact, civil servants go through the same kind of exercise that judges engage in when they have to deliberate and then reason a judgment. So, they may intuitively have some kind of idea of why they want to adopt an algorithm and what it should look like. But if they have to start answering questions about that, have to discuss this, and have to write down their answers, their intuitive choice somehow may not look that convincing anymore. That may trigger further discussion and it may even lead to a different outcome. Indeed, completing FRAIA is only for a small part about writing down answers; it really is about the discussion in a team and the thought and reasoning processes this triggers. That really convinces people to work with it, and it creates accountability to the extent that the team involved can convincingly explain why they have made a certain choice.

But, indeed, it could be that FRAIA is working so well because it is not obligatory yet, and the bodies working with it are intrinsically motivated to work on responsible AI. If it was mandatory to use FRAIA in a great many cases, this might lead to some fatigue and FRAIA could, in the end, still become a paper tiger.

VM: *FRAIA clearly underscores the importance of the impact of algorithmic systems on fundamental rights. Please tell us more about the function of the Fundamental Rights Roadmap in part 4 of FRAIA. Could you elaborate on how it helps to reconcile potential value conflicts?*

JG: The function of the Fundamental Rights Roadmap is to identify what kind of fundamental rights risks there are in relation to the algorithm the team wants to work with, how serious these risks are, and whether there could be a justification for introducing that algorithm regardless of these risks.

The first step—making an inventory of which fundamental rights could be affected by an algorithm—is already a revelation to some civil servants. People are usually aware of privacy and data protection issues, but they tend to easily overlook other fundamental rights, such as procedural and non-discrimination rights. For that reason, we made this long list of fundamental rights so that civil servants can relatively easily look up which rights could be affected by "their" algorithm. Subsequently, they have to go through the other steps of the roadmap for each of these rights. That means they will have to identify whether there is specific legislation that applies with respect to the considered fundamental right (think of the GDPR or of equal treatment legislation); if so, they have to apply that. If there isn't, they will have to estimate the seriousness of the infringement of that fundamental right. Then, considering the aims of introducing a specific algorithm that had been defined in an earlier part of FRAIA, and taking into account the seriousness of the expected infringement, the team will have to look into the effectiveness and necessity of the algorithm. Will the chosen algorithm really help to realize the objectives, or will it be a limited contribution? Are there alternative policy instruments available that would interfere with the fundamental rights to a lesser extent, or are there mitigation strategies conceivable? If so, then they would typically be preferable.

If the algorithm "survives" all these steps, the team will arrive at the last, and most difficult, step, which requires a balance of rights and interests. Metaphorically speaking, this means that they put the fundamental rights that will be affected and the extent to which they will be affected on one scale of the balance, and on the other scale there will be the objectives of public policy, their importance, and the extent to which they can be realized by the algorithm. The team will then have to compare these, and, in the end, make a choice, or leave the choice to be made by politically responsible actors. Either way, we advise teams to write down as carefully as possible why a particular choice was made and why they believe that the benefits

would outweigh the costs of interfering with fundamental rights. If they have done so, and a decision based on the algorithmic system ever comes before the higher authority, they will at least have a convincing story to tell.

Of course, all of this does not guarantee a substantively good outcome, but it is the best we can do in terms of making the process such that it guarantees an outcome that is acceptable to most people.

VM: *What are the main benefits and challenges for governmental organizations in implementing FRAIA? Do you think it could be exported and used in other countries?*

JG: The instrument helps civil servants in making good policy and I think most of them appreciate that. By going through FRAIA, they can explain why they made certain choices and if a mistake was made, they can trace back exactly where in the decision-making process things have gone wrong. This helps them to feel safe and certain about making a decision. It also guarantees some intersubjectivity because it is always a team exercise. All team members should agree on the outcome of FRAIA or at least they should be able to agree to disagree about it. However, one thing I am not sure about is whether and to what extent this is a cultural matter. FRAIA might work in the Dutch context, which seems to be quite open and non-hierarchical, but I am not sure if this is the same in different cultures of governance. That needs checking and perhaps some experimentation.

Moreover, if we were to get the chance, I think we would make some additional changes. One of the issues that has come up is the need for further alignment of FRAIA with the data protection impact assessments (DPIAs), as having to do both might be a real burden on organizations. Additionally, there has been a question of whether a "quick scan" could be developed to see if an algorithm would have a real impact on fundamental rights, so public bodies would only need to use FRAIA in such cases. Finally, in a way, enforcement could be a challenge. So far, FRAIA is not obligatory, but there are some ideas of making it mandatory. If that were to happen, you would also have to think about how to implement, monitor, and audit it. However, I am not sure if the instrument lends itself well to mandatory application and if it still would work well then.

VM: *FRAIA is designed to be implemented by governmental organizations. Do you see a role for an instrument like FRAIA to be extended to the commercial sector? In this respect, how does FRAIA relate to the EU Artificial Intelligence Act (AI Act)?*

JG: Yes, I think many of the questions we ask governmental organizations are equally relevant to private organizations. We would probably need to make some changes because public values do not necessarily need to be the main consideration for many companies, and we need to see to what extent it is possible to translate this particular model to private organizations.

Indeed, in developing FRAIA, we looked into public governance theory and studies and what we know about how public authorities and civil servants behave. Perhaps it would be possible to do something similar together with scholars in the field of how companies behave and what kind of incentives they have. It would be great if there would be ways to apply this without just having to rely on the "stick approach" of imposing FRAIA as a full-blown obligation, but to be able to find some kind of "carrot"—an incentive that makes it actually attractive for private organizations to engage in a FRAIA-like exercise.

When it comes to the AI Act, everyone is currently waiting for the final text to be released. This act includes a fundamental rights impact assessment, which is actually based on our previous work. GroenLinks (Green Party) MEP Kim Sparrentak knew about our FRAIA and did a great deal of fantastic lobbying to ensure that the European Parliament was in favor of adding some kind of an impact assessment to the AI Act. It is not yet known if the EU will also provide a kind of blueprint for the impact assessment that we have to apply. We will have to wait and see how this unfolds and if we can still opt to use FRAIA for the assessment process.

VM: *Are there any emerging AI trends or developments* (*e.g., the rise of generative AI applications*) *that could affect the future relevance of this instrument? If so, how might it be affected?*

JG: Taking the example of the new large language models (such as ChatGPT) or distributed ledger technologies (such as non-fungible tokens), I think FRAIA can be applied to them, too. It is clear that the impact is probably going to be bigger, and that the data used are different, but the questions contained in FRAIA are equally relevant and will remain more or less the same. That said, I am wondering whether it could be useful to have more dedicated and precise questions or modules that are more closely geared to specific technologies. In fact, I think there is still a lot of work to be done to further develop FRAIA, and I really hope will be given the opportunity to do so.

References

DEDA. N.d. "Data Ethics Decision Aid." https://deda.dataschool.nl/en/.

Gerards, Janneke, Mirko Tobias Schäfer, Arthur Vankan, and Iris Muis. 2022. *Impact Assessment: Fundamental Rights and Algorithms*. The Hague: Ministry of the Interior and Kingdom Relations. https://www.government.nl/documents/reports/2022/03/31/impact-assessment-fundamental-rights-and-algorithms.

Google. 2023. "Privacy Policy." November 15. https://policies.google.com/privacy.

Nederland Digitaal. 2021. "The Dutch Digitalisation Strategy 2021." https://www.nederlanddigitaal.nl/documenten/publicaties/2021/06/22/the-dutch-digitalisation-strategy-2021-eng.

Reisman, Dillon, Jason Schultz, Kate Crawford, and Meredith Whittaker. 2018. "Algorithmic Impact Assessments: A Practical Framework for Public Agency." *AI Now*, October 4. https://www.nist.gov/system/files/documents/2021/10/04/aiareport2018.pdf.

Van Til, Gijs. 2019. "Netherlands: Automating Society 2019." *Algorithm Watch*. https://algorithmwatch.org/en/automating-society-2019/netherlands/.

Wieringa, Maranke. 2023. "Hey SyRI, Tell Me about Algorithmic Accountability: Lessons from a Landmark Case." *Data & Policy* 5: e2. https://doi.org/10.1017/dap.2022.39.

About the Author

Viktorija Morozovaite is assistant professor at the University of Amsterdam.