

## Transparante AI: van wettekst naar de praktijk

mr. Ljubiša Metikoš, promovendus

*Universiteit van Amsterdam, Faculteit der Rechtsgeleerdheid*

*Instituut voor Informatierecht, Paul Scholten Centre for Jurisprudence, RPA Human(e)AI*

Het is misschien een platitudo om te stellen dat AI ondoorzichtig kan zijn. Toch lijkt het belang van transparantie en uitlegbaarheid nog steeds niet doorgedrongen te zijn binnen de Nederlandse overheid. In de praktijk zien we namelijk dat verschillende overheidsinstanties keer op keer ondoorzichtige AI systemen inzetten, die burgers profileren en discrimineren.<sup>1</sup> In deze korte primer voor mijn presentatie tijdens de Jonge VAR, ga ik in op een paar van de obstakels voor transparant AI-gebruik binnen de Nederlandse overheid. Daarbij benadruk ik ook hoe het gebruik van AI transparantie juist kan helpen waarborgen. Ten slotte sluit ik af met een korte introductie op de transparantieplichtingen die bestaan binnen het informatierecht en geef ik een paar aanwijzingen over hoe de Nederlandse overheid aan deze verplichtingen kan voldoen.

Om deze primer te starten is het wellicht handig om eerst te kijken naar een inmiddels welbekend voorbeeld van ondoorzichtig AI-gebruik binnen de Nederlandse overheid: de toeslagenaffaire. Jarenlang gebruikte de belastingdienst een AI systeem om de kans op fraude onder burgers te voorspellen. Dit systeem werd in het geheim ingezet en selecteerde burgers die een dubbele nationaliteit hadden, donaties aan moskeeën maakten, van Surinaamse of Antilliaanse komaf waren etc. Het gebruik van dit systeem werd stelselmatig verzwegen tijdens de rechtszaken van de gedupeerden, in de beantwoording van Kamervragen en ook in het reageren op Wob/WOO-verzoeken.<sup>2</sup>

De toeslagenaffaire laat goed zien dat er grote organisatorische obstakels voor de transparantie van AI kunnen ontstaan wanneer een overheidsorganisatie discriminatie of ander onrecht wil verhullen.<sup>3</sup> Maar dat is niet de enige hindernis voor het transparant maken van AI systemen. Ook de bedrijven die deze systemen helpen ontwikkelen kunnen een blokkerende rol hebben. Het kan zo zijn dat deze ontwikkelaars zich beroepen op hun auteursrechten of handelsgeheimen om te voorkomen dat er informatie uitlekt over het systeem die hun concurrentiepositie kan schaden.<sup>4</sup> Verder kunnen overheidsinstanties zich ook zorgen maken over de fraudegevoeligheid van deze systemen, als ze blootgesteld worden aan de buitenwereld.<sup>5</sup>

In het geval van complexe ‘black-box’ systemen, zoals AI modellen die gebruik maken van ‘neural networks’,<sup>6</sup> kunnen óók *technische* obstakels het onmogelijk maken om een systeem te begrijpen. Hierdoor kan zelfs de overheidsinstantie in kwestie het overzicht kwijt raken over hoe het systeem werkt. In de praktijk kunnen burgers daardoor geconfronteerd worden met, bijvoorbeeld, een risico-indicatiescore die, zonder enige uitleg over hoe het tot stand is gekomen, klakkeloos voor waar wordt aangenomen.<sup>7</sup> Zulke systemen verwarren helaas vaak correlatieve en causatieve

---

<sup>1</sup> Parlementaire ondervragingscommissie Kinderopvangtoeslag, ‘Ongekend Onrecht’, 17 december 2020; Salwa van der Gaag, ‘DUO negeerde signalen over etnisch profileren bij fraudeonderzoek’, NOS 17 juli 2023.

<sup>2</sup> Parlementaire ondervragingscommissie Kinderopvangtoeslag, ‘Ongekend Onrecht’, 17 december 2020, p. 8.

<sup>3</sup> Ari Waldman, ‘Power, Process, and Automated Decision-Making’, *Fordham Law Review* 88, nr. 2 (1 november 2019): 613.

<sup>4</sup> Paul B. de Laat, ‘Algorithmic Decision-Making Employing Profiling: Will Trade Secrecy Protection Render the Right to Explanation Toothless?’, *Ethics and Information Technology* 24, nr. 2 (5 april 2022): 17, <https://doi.org/10.1007/s10676-022-09642-1>.

<sup>5</sup> Dan L. Burk, ‘Algorithmic Legal Metrics’, *The Notre Dame Law Review* 96, nr. 3 (2021): 1147-1203.

<sup>6</sup> Bartosz Brożek e.a., ‘The Black Box Problem Revisited. Real and Imaginary Challenges for Automated Legal Decision Making’, *Artificial Intelligence and Law* 32, nr. 2 (1 juni 2024): 427-40, <https://doi.org/10.1007/s10506-023-09356-9>.

<sup>7</sup> Karen Yeung en Adam Harkens, ‘How do “technical” design choices made when building algorithmic decision-making tools for criminal justice authorities create constitutional dangers? (Part I)’, *Public Law* 2023, nr. Apr (30 april 2023): 265-86.

verbanden met elkaar.<sup>8</sup> Toch worden deze onwetenschappelijke systemen dikwijls gezien als waarzeggers, die op wiskundige wijze de toekomst feilloos zouden kunnen voorspellen.<sup>9</sup>

Dit wil niet zeggen dat een AI-ondersteunde overheid per definitie altijd ondoorzichtig hoeft te zijn. AI systemen hebben namelijk de potentie om véél transparanter te zijn dan menselijke besluitnemers. Ambtenaren kunnen immers net zo goed geleid worden door vooringenomenheid en andere belangen.<sup>10</sup> Het is daarnaast ook niet gemakkelijk om te achterhalen welke ambtenaar discriminerende of foute besluiten heeft genomen. Sommige typen AI systemen kunnen echter van begin tot eind gevolgd worden in hoe ze een bepaalde input transformeren tot een bepaalde output.<sup>11</sup> Hiermee kan transparantie juist verbeterd worden binnen de overheid.

Wel is het zo dat het waarborgen van transparantie, o.a. met als doel het mogelijk maken van weerspraak en kritiek, sterk afhankelijk is van zowel technische als organisatorische maatregelen. Enerzijds moet er de (politieke) wil zijn om een gezonde en empathische bestuurscultuur vorm te geven binnen elk niveau van de ambtenarij, van de ambtelijke top tot de 'street-level bureaucrat'.<sup>12</sup> Daarnaast moeten er ook de juiste technische maatregelen genomen worden in overleg met de ontwikkelaars van deze systemen, om transparantie al vanaf het begin van het ontwerpproces te waarborgen.

Deze maatregelen zijn verder ook nodig om te voldoen aan het brede scala aan transparantieverplichtingen die volgen uit de AVG en de AI Wet. Tijdens mijn presentatie zal ik met name ingaan op een paar van de belangrijkste transparantieverplichtingen die in deze wetten te vinden zijn. Ik zal o.a. focussen op het verbod op volledig geautomatiseerde besluitvorming (art. 22 AVG), het recht op een uitleg (art. 13, 14, 15 AVG en art. 86 AI Wet), de verplichting tot menselijk toezicht (art. 22 AVG en art. 13 en 14 AI Wet) en het Europees hoog-risico algoritmeregister (art. 49 AI Wet).

Tijdens mijn presentatie zal ik kort ingaan op wat de technische en organisatorische vereisten zijn die voortvloeien uit deze wetsbepalingen. Ik zal laten zien hoe overheidsinstanties bijvoorbeeld kunnen putten uit de ontwerpprincipes en methodologieën die worden voorgesteld op het gebied van XAI (explainable AI). Verder stel ik ook dat bepaalde ethische principes en doelstellingen nauwer betrokken kunnen worden in het ontwerpproces van AI systemen. Ik put daarvoor met name uit mijn eerdere werk op het gebied van AI en interdisciplinair juridisch onderzoek en het concept van 'contestability by design'.<sup>13</sup>

Concluderend hoop ik dat deze primer heeft laten zien dat het noodzakelijk is om verschillende transparantie maatregelen *voorafgaand* aan de inzet van AI te nemen. Mijn presentatie roept u, de lezer, dan ook op om nog verder te reflecteren over hoe AI ingezet kan worden op een wijze die de fundamentele waarde van transparantie kan respecteren.

---

<sup>8</sup> Yeung en Harkens.

<sup>9</sup> Daniel McQuillan, 'Data Science as Machinic Neoplatonism', *Philosophy & Technology* 31, nr. 2 (juni 2018): 253-72.

<sup>10</sup> John Zerilli e.a., 'Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard?', *Philosophy & Technology* 32, nr. 4 (1 december 2019): 661-83, <https://doi.org/10.1007/s13347-018-0330-6>.

<sup>11</sup> Cynthia Rudin, 'Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead', *Nature Machine Intelligence* 1, nr. 5 (mei 2019): 206-15, <https://doi.org/10.1038/s42256-019-0048-x>.

<sup>12</sup> Sofia Ranchordas, 'Empathy in the Digital Administrative State Automating the Administrative State', *Duke Law Journal* 71, nr. 6 (2022 2021): 1341-90.

<sup>13</sup> Ljubiša Metikoš, 'An Interdisciplinary Toolbox for Researching the AI-Act', *Verfassungsblog*, 8 september 2023, <https://doi.org/10.17176/20230908-062850-0>; Ljubiša Metikoš, 'The AI Act: Weak, Weaker, Weakest', *Mediaforum* 3 (2024): 73-74; Ljubiša Metikoš, 'Explaining and Contesting Judicial Profiling Systems', *Technology and Regulation* 2024 (13 september 2024): 188-208, <https://doi.org/10.26116/techreg.2024.017>.